

Statistics of Fuzzy Data

HIT

数学·统计学系列

模糊数据统计学

王忠玉 吴柏林 著

哈尔滨工业大学出版社
HARBIN INSTITUTE OF TECHNOLOGY PRESS



数学·统计学系列

模糊数据统计学

王忠玉 吴柏林 著

哈尔滨工业大学出版社
HARBIN INSTITUTE OF TECHNOLOGY PRESS

内容提要

本书把可利用模糊数据刻画的模糊现象纳入到统计学的研究范畴，将模糊数学的研究思想及方法与统计学融合，并且结合软计算方法，提出模糊数据描述统计量、模糊数据问卷调查、模糊数据均数估计、模糊数据假设检验、模糊数据回归模型、模糊数据样本排序及非参数检验方法、模糊数据时间数列分析及预测等，阐述深入浅出，理论结合实际，在探讨模糊统计基本理论基础上，构建一套完整的专门研究模糊数据的统计方法体系，反映出发展中的模糊数据统计学的最新成果及趋势，为研究和处理复杂性数据提供一种新的软计算方法及理论。

本书可作为数学、应用数学、统计学等专业的本科高年级的前沿专题及研究生教材，也可作为数量经济学、金融学、心理学、社会学等专业硕士研究生以及对模糊统计感兴趣的科技人员的有关教材及参考书。

图书在版编目(CIP)数据

模糊数据统计学/王忠玉著．—哈尔滨：哈尔滨工业大学出版社，2008.3

ISBN 978-7-5603-2657-3

Ⅰ.模…　Ⅱ.王…　Ⅲ.模糊数学-统计学　Ⅳ.C8

中国版本图书馆 CIP 数据核字(2008)第 016199 号

策划编辑　刘培杰
责任编辑　王勇钢
封面设计　卞秉利
出版发行　哈尔滨工业大学出版社
社　　址　哈尔滨市南岗区复华四道街 10 号　邮编 150006
传　　真　0451-86414749
网　　址　http://hitpress.hit.edu.cn
印　　刷　黑龙江省教育厅印刷厂
开　　本　787mm×1092mm　1/16　印张 15.25　字数 266 千字
版　　次　2008 年 3 月第 1 版　2008 年 3 月第 1 次印刷
书　　号　ISBN 978-7-5603-2657-3
印　　数　1～2 000
定　　价　48.00 元

(如因印装质量问题影响阅读，我社负责调换)

前　言

统计是用来分析、处理自然科学及社会科学信息的工具，在复杂的自然或社会现象中，人们可以借助于样本数据所提供的信息，经过归纳分析、推断检验、决策、预测等过程，使我们对现实状况更了解、更好地处理现实世界的问题。传统统计学的目的主要针对各类信息，拟定一套对数据进行分析、估计、检验的测量方法，其过程包括：(1)设定合适的理论或模型；(2)收集样本数据(包括实验设计、抽样或模拟)；(3)进行数据分析；(4)实施推断估计；(5)假设检验；(6)决策或预测。

近年来，由于经济全球化、科学技术快速发展、计算机人工智能科技发展一日千里，经济社会现象日趋多变、全球信息网络化，以往的社会科学研究大多利用传统的统计分析方法，如今却因为时代的不断进步，而逐渐不符合现今多变环境的复杂性，以致常常感到研究方法的缺乏与不适用。传统的统计分析工具之所以不敷应用，一个主要的问题是：如何更有效处理分析日益复杂、海量的网络情报数据。虽然数据挖掘的兴起及发展，解决了不少数据分析的问题，但是，对于如何处理非实数样本数据，比如区间数据、多值数据形式的模糊数据样本，应用建立在实变函数与概率论基础上的传统统计方法已无法有效地进行分析和预测。尤其是，我们在决策过程中所遇到的不确定性问题比我们想象的更为复杂。情报信息除了随机性外，还包括不完全的信息、部分的已知知识，或者对环境模糊的描述等。

此外，从人类自身发展对外部现实世界事物认识的角度来看，如何对人的感知认识(绝大多数认识都是模糊性的观念等)——进行数量化，并加以统计信息处理及运算，进而实施统计推断及预测等一直困扰着统计学研究者。

事实上，我们所获得的信息源自测量与感知，而感知信息中的不确定因素，主要是我们的语言对某些概念表达模糊所引起的。显然，要作出比较好的判断，我们必须尽量将所能得到的信息都考虑在内。这包括用自然语言描述的行为、意义等的属性信息。因此，我们需要利用概率将模糊概念数学模式化，其实这也展示不确定性的另一种形式。模糊理论是一种定量化处理人类语言、思维的一门新兴学科。模糊逻辑并非如字面意思那样的马虎、不精确，而是面对生活上各种各样的不确定性，以更合理的规则去分析及管理控制，以期得到更有效率、更合乎人性与智慧的结果。模糊数据统计并不模糊，它是处理不确定事

件的一种新技术，带领我们从古典的统计估计与检验研究计算，进入一个需要软计算、稳健性的高科技的 e 时代。

在传统的统计推断方法中，为了了解未知总体参数值，我们经常借助于一些评估准则，找出适当的统计量来对总体参数进行估计。平均数是认识总体集中趋势最重要的总体参数之一，我们常常利用无偏估计量，比如样本平均数来估计总体集中趋势。然而，在日常生活中，总体平均数经常是带有模糊、不确定性的语意变量，或是一个可能区间，传统的估计量评估准则及估计方法便无法适用于此种情形。

在当今信息化时代，一个基本的问题是如何管理把不确定性（包括不准确性与模糊性）考虑进去的认知过程（Cognitive Process），这无论是在技术、经济学、工程、生物医学等方面，还是在实践上都是一个意义重大的问题。实际上，现实情形是促使人们研究这类问题的最初动机来源。

从理论观点来看，信息与不确定性是紧密联系的两个概念。事实上，不确定性源于信息的缺乏。另一方面，可以把信息看成是关于不确定性的缩减，尽管此观点很重要，但这只是考察信息概念的一种可行方式。因此，信息领域中的任何概念上的丰富充实都通过对新的不确定性形式的需求管理而与之对应。

在最近的二三十年里，不确定性理论在其概念领域及工具方法上都取得了显著的扩展。而这些扩展与数学思想上的两个扩展相互促进影响：一方面，经典的可加性测度理论向单调而非可加性测度（比如，可能性测度、信念函数、区间值可能性）的扩展；另一方面，经典集合理论以其标准形式和非标准形式向模糊集合理论的扩展。

统计推理是在部分信息条件下进行推理的方式，因此，也是在不确定性存在条件下实施推理的方式。上述曾提及的信息与不确定性领域的一些发展自然而然地对统计科学产生影响。众所周知，不确定性是决策分析研究中的最主要的困难所在。事件的不确定性主要有两种表现形式：随机性和模糊性。

自从美国控制论专家、著名模糊数学家扎德教授（L. A. Zadeh）在 1965 年提出模糊集合理论以来，最初的模糊集合理论在逻辑学、数学、技术及工程等方面得到了迅速普及与推广，经过四十多年的不断研究及发展，模糊数学方法在经济管理、工程、决策、聚类分析、控制等诸多领域，不论在理论研究，还是实际应用中均取得了相当丰硕的成果。对于统计方法来说，这提供了富有成果的新思想和新工具。

模糊数学方法应用到统计分析中，特别是近十多年来取得了长足的发展，目前这类研究正在蓬勃的发展中。

本书基于模糊数学的研究思想及方法，并结合软计算方法，定义出模糊数

据样本均数、模糊数据样本众数以及模糊数据中位数等，然后给出它们的许多相关性质。同时，针对模糊数据的参数估计量，我们提出适当而可行估计法的评判准则。对于古典的统计检验，必须陈列明确的假设。当我们要想检验两总体平均数是否有差异时，零假设（或称为原假设）是“两个平均数相等”。然而，有时我们想要知道的只是两平均值是否模糊相等，此时传统的检验方法并不适用于这种包含不确定性的模糊数据假设检验。因此，本书提出基于模糊数据样本的统计检验方法，针对模糊均数相等、模糊属于以及卡方同质性检验作出进一步的探讨。

本书内容自给自足，第 1 章和第 2 章是全书的基础内容，介绍模糊数学的基本知识，并阐述模糊数据的一些软计算方法等。而从第 3 章至第 10 章为本书的重点内容——模糊数据统计学，内容涉及模糊数据问卷调查，模糊聚类分析，模糊数据回归分析，模糊数据非参数统计，模糊数据时间数列分析及预测。我们列举出社会科学的一些应用实例，比如模糊问卷调查表，乐观测量表，风景区满意度调查，茶叶模糊分类，模糊回归及景气循环，模糊时间数列。第 11 章内容涉及当今模糊数据统计学发展中所遇到的问题与前景。

本书内容新颖，阐述深入浅出，理论结合实际，反映出发展中的模糊数据统计学的最新成果及趋势，为社会科学中的数量化方法及理论提供一种新颖的软计算方法及理论。期望以此抛砖引玉，开创 21 世纪模糊数据统计理论及应用研究的崭新领域。

全书内容写作安排如下：王忠玉撰写第 1 章、第 2 章、第 5 章、第 6 章、第 9 章以及附录 1，其余内容由吴柏林撰写。

王忠玉于哈尔滨
吴柏林于中国台北
2007 年，秋季

目　　录

第 1 章　引论 …… 1

1.1　人的感知与模糊性 …… 1

1.2　模糊集合与隶属度 …… 3

1.3　模糊集合运算 …… 6

1.4　模糊矩阵 …… 10

1.5　截集与支集 …… 12

第 2 章　隶属函数及软计算方法 …… 18

2.1　隶属函数 …… 18

2.2　模糊数据与软计算 …… 25

2.3　模糊关系 …… 33

2.4　模糊等价关系 …… 37

2.5　语意计量与相似度 …… 39

第 3 章　模糊数据描述统计量 …… 44

3.1　模糊数据样本均数 …… 45

3.2　模糊数据样本众数 …… 47

3.3　模糊数据样本中位数 …… 51

3.4　模糊数据统计量的一些性质 …… 54

第 4 章　模糊数据问卷调查 …… 60

4.1　社会思维的分歧性与模糊性 …… 60

4.2　模糊数据问卷设计与特征获取 …… 62

4.3　模糊数据测量表 …… 65

4.4　个案研究:选民投票意向与选情预测 …… 75

4.5　结论 …… 87

附表 4.1　乐观测量表 …… 88

附表 4.2　ABC 市长选民投票意向问卷 …… 89

第 5 章　模糊数据均数估计 …… 91

5.1　模糊数据总体均数 …… 91

5.2　模糊数据总体均数最佳估计方法 …… 97

5.3 模糊数据估计量的评判准则 …… 102
第 6 章 模糊数据假设检验 …… 104
6.1 距离 …… 104
6.2 模糊数据总体均数检验 …… 107
6.3 模糊分类数据的卡方同质性检验 …… 110
第 7 章 模糊聚类分析 …… 113
7.1 模糊聚类法 …… 113
7.2 模糊权重分析与判定程序 …… 121
7.3 加权模糊分类 …… 126
7.4 茶叶等级分类实例 …… 128
7.5 结论 …… 132
附表 7.1 中国台湾 69 个茶树品种春季制茶质量的评分比较 …… 133
附表 7.2 中国台湾 69 个茶树品种全年制茶质量的评分比较 …… 134
附表 7.3 中国台湾 69 个茶树品种春茶的分类结果 …… 135
附表 7.4 中国台湾 69 个茶树品种全年的分类结果 …… 136
附表 7.5 中国台湾 69 个茶树品种春茶制茶质量加权模糊聚类分析 …… 137
附表 7.6 中国台湾 69 个茶树品种全年制茶质量加权模糊聚类分析 …… 137
第 8 章 模糊数据回归模型及应用 …… 138
8.1 模糊数据回归简介 …… 138
8.2 模糊数据回归建构 …… 139
8.3 模糊数据回归的参数估计 …… 140
8.4 景气对策信号实例 …… 145
8.5 篮球比赛攻防策略实例 …… 148
8.6 结论 …… 153
第 9 章 模糊数据样本排序及非参数检验方法 …… 155
9.1 模糊数据样本排序 …… 155
9.2 模糊数据中位数的符号检验及应用 …… 158
9.3 模糊数据样本排序方法应用于威尔克森符号秩检验 …… 160
9.4 模糊数据样本排序方法应用于威尔克森秩和检验 …… 162
9.5 模糊数据样本排序方法应用于克鲁斯考尔和沃利斯秩检验 …… 164
9.6 结论 …… 165

第 10 章　模糊数据时间数列分析及预测 …… 167

10.1　引言 …… 167

10.2　模糊数据时间数列的统计分析 …… 172

10.3　模糊数据自回归时间数列($FAR(p)$) …… 174

10.4　模糊数据时间数列模型建立 …… 177

10.5　景气对策信号实例 …… 185

10.6　结论 …… 192

第 11 章　问题及展望 …… 194

附录 …… 199

附录 1　偏大与偏小型隶属函数表 …… 199

附录 2　二项分布累积概率 $P(X \leqslant x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}$ …… 203

附录 3　标准正态分布累积概念：$P(Z \leqslant Z_\alpha) = \int_{-\infty}^{Z_\alpha} f(z)\mathrm{d}z = 1 - \alpha$ …… 207

附录 4　t 分布的临界值：$t_\alpha(v) P(T \geqslant t_\alpha(v)) = \int_{t_\alpha(v)}^{\infty} f(x;v)\mathrm{d}x = \alpha$ …… 208

附录 5　卡方分布的临界值：$\chi_\alpha^2(v) P(\chi^2 \geqslant \chi_\alpha^2(v)) = \int_{\chi_\alpha^2(v)}^{\infty} f(x;v)\mathrm{d}x = \alpha$ …… 209

附录 6　F 分布的临界值：$F_\alpha(v_1, v_2) P(F \geqslant F_\alpha(v_1, v_2)) = \int_{F_\alpha(v_1, v_2)}^{\infty} f(x; v_1, v_2)\mathrm{d}x = \alpha$ …… 211

附录 7　威尔克森符号秩检验概率 …… 217

附录 8　威尔克森秩和检验概率 …… 218

参考文献 …… 219

第 1 章　引　　论

第 1 章和第 2 章是全书的内容基础。第 1 章以人对事物现象认识的感知与模糊性作为切入点，阐述模糊性是人对事物认识的一种表征及反应。然后，引入模糊集合的定义及刻画模糊集合的表示函数 —— 隶属度(Membership)，对模糊集合的各种运算、模糊矩阵、模糊截集以及支集等内容进行较详细的讨论，并以各种事例阐明概念和运算。

对于已经学习过模糊数学的人来说，可以直接从第 3 章模糊数据描述统计量开始阅读及学习。对于那些没有模糊数学知识的人来说，第 1 章和第 2 章的内容则是极为基本而必需的，因为它们对学习后续各章的知识提供刻画与表述模糊现象的认识框架及数量化描述方法。本书内容写作的自给自足特点正是为以前没有学习过模糊数学知识的人而设计的。

1.1　人的感知与模糊性

人类的思维主要是源于对自然现象和社会现象的认知意识，而人类的知识语言也会因本身的主观意识、时间、环境以及分析事情的角度不同而具有模糊性。模糊理论的产生即是参考人类思维方式对环境所用的模糊测度与分类原理，给予较稳健的描述方式，以便处理多元复杂的暧昧关系与不确定性现象。

通常，人类思维有两种形式，一种为形式化思维(Formal Thinking)，另一种为模糊思维(Fuzzy Thinking)；前者具有逻辑性和顺序性的思考，而后者则是全体性和综合性的思考。当面临决策判断而进行思考时，基于形式化思维的二元逻辑，常常很难表示出人类思考的多元逻辑特性。

当有人说，他今天感到很快乐时，究竟他对于快乐的认知为何呢?什么样的测量标准可以称得上快乐呢?或是这样的感觉持续多久的时间以上才能算是快乐呢?然而，对这样的问题，每个人的回答皆因其主观性而有所不同，即使回答者为同一人，也会因为所处的环境、或是外在条件的不同，而可能出现与之前相异的答案。诸如此类很多的论点和问题，都不是能够用绝对的二元逻辑所可以界定的。其原因则皆源自人类思维的模糊性。但是，人们却经常被要求作出绝对

的判断或选择，以人性的观点来看，这是十分不合理的。

对不确定性的事物作出决策，是相当重要的人类活动。不确定性是决策分析研究中的最主要的困难所在。事件的不确定性主要有两种表现形式：随机性与模糊性。

如果不确定性仅仅是由于事物的随机所引起的，那么模糊数据统计分析的发展则为这类决策活动提供了很好的理论依据。事实上，我们在决策过程中所遇到的不确定性问题，往往不只是由于事物的随机所引起，这种不确定性还可能是：不完全的信息、已知的部分知识、对环境模糊的描述等，这类信息出自于测量与感知中的不确定因素，主要是由于我们的语言及人类思维对某些概念表达模糊所引起。通常，这些不明确性比我们想象的要复杂得多。

显然，如果要对人类思维的模糊性作出比较好的判断，我们必须尽量将所得到的信息都考虑在内，特别是属性问题。由于属性问题本身的模糊性与不精确性，若我们利用此假设的精确值来作出因果分析与计量度量，可能造成判定偏差及决策失误，甚至会扩大预测结果与实际状态之间的差异。因此，对于这些在思考认知不易表达完善的属性问题，借助于软计算方法与模糊数据统计分析可以更明确地表达出来。

有些学者认为，模糊理论是研究不确定的现象，应该与概率论相类似。然而，概率论是研究随机性问题，随机性虽不确定，但那是因为条件不充分引起的，事件的发生是随机的，事件发生之后就是确定的。例如，掷一个公正骰子，出现 1,2,3,4,5,6 点之概率均为 1/6，当掷完一次骰子之后，出现多少就是多少。而模糊理论的事件本身却是模糊的、不精确的，例如，回答某家庭经济状况属于不错、小康或中等等，这些均不属于随机，而是事件本身的不完整性与不精确性。扎德(L. A. Zadeh, 1999) 更建议引用感知测度(Perception Measure) 和软计算(Soft Computing) 共同作为模糊函数估计量，这种应用模糊概念将属性关系数学模型化的方法，我们统称为软计算方法。实际上，软计算是一些计算方法的综合，其核心包括模糊逻辑、神经元算法、遗传算法以及概率算法。这些方法之间相互补充、协同作用而无优劣之分。

模糊理论的概念，主要强调人的喜好程度不需非常清晰或数值精确，因此，对人的感知而言，模糊模型比直接指定单一物体一个值，更适合于评估人所认识的物体之间多元或相关特性。

1.2 模糊集合与隶属度

普通集合对应于二值逻辑，表现为布尔代数。元素 x 属于或不属于某一个集合 A，非常清楚，毫不含糊。古典集合论在数理科学上已经建立起一套相当完善、系统的逻辑体系。

但是，若将此集合关系应用于描述某些事物现象时，经常会出现不合理的情况。因为某些现象并不一定存在"非此即彼"的关系。例如，进行某一教学单元后，将班级的学生划分成"精熟"和"不精熟"两类，这样的划分很明显的有不合理之处，因为学生的精熟程度并非是二元的现象，而是有各种不同精熟程度连续性的特性。自扎德(L.A.Zadeh，1965)提出模糊集合及模糊理论以来，此种思维可以解释许多事物现象。模糊理论将元素和集合之间的关系，通过介于[0，1]之间的隶属度加以描述。

设 U 表示所要研究对象的集合，称为论域。对于论域 U 的一个子集 A，可以利用它的特征函数(Characteristic Function)来表示。令

$$I_\mu(x)=\begin{cases}1 & x\in A\\ 0 & x\notin A\end{cases} \tag{1.1}$$

$I_\mu(x)$ 表示在 U 上定义的取值于 $\{1,0\}$ 的函数，称为集合 A 的特征函数，$I_\mu(x)$ 明确表示了集合 A，对于 $x\in A$，若 $I_\mu(x)=1$，则表示 x 是 A 中的元素；若 $I_\mu(x)=0$，则表示 x 不是 A 中的元素。由此，下面给出模糊集合的定义。

【定义 1.1】 模糊集合

设 U 表示一个论域，U 上的一个模糊子集 A 是指对于任何 $u\in A$ 都有一个实值函数 $\mu_A(u)\in[0,1]$ 与之对应，也就是

$$\mu_A: U\to[0,1]$$
$$u\mapsto \mu_A(u)\in[0,1] \tag{1.2}$$

映射 μ_A 称为 A 的隶属函数，$\mu_A(u)$ 表示元素 u 属于 A 的程度，称为 u 对 A 的隶属度。

U 上的模糊子集所构成的集合全体称为模糊幂集，记为 $\Phi(U)$，即

$$\Phi(U)=\{A\mid \mu_A: U\to[0,1]\} \tag{1.3}$$

显然，$\Phi(U)\supset \Pi(U)$，其中 $\Pi(U)$ 表示普通幂集。

模糊子集是由其隶属函数来决定的。隶属函数则是普通集合中特征函数的推广。一般地说，模糊集合 A 是一个抽象的事物，而隶属函数 μ_A 则是一个具体

的。通常，只能通过 μ_A 来认识及掌握模糊集合 A。

隶属函数是模糊理论的基础，它是从传统集合中的特征函数所衍生出来的，用以表达元素对模糊集合的隶属度，其范围介于 0 到 1 之间。对于元素和集合的关系，古典集合将元素和集合之间的关系用特征函数来说明，亦即若 $u \in A$，则 $I(u)=1$；若 $u \notin A$，则 $I(x)=0$。但是，扎德(L. A. Zadeh，1965)在模糊集合论中提到，若一个元素属于某一个集合的程度越大，则其隶属度值越接近于 1，反之则越接近于 0。

【例 1.1】 天气好坏的模糊集合。

设论域 $U=\{$星期一，星期二，星期三，星期四，星期五，星期六，星期日$\}$。若从星期一到星期三是好天气，而从星期四到星期日都是坏天气，按普通集合观点，特征函数有 $f_A(u)=1, u$(好天气)$\in A$；$f_A(u)=0, u$(坏天气)$\notin A$。隶属度分别为：μ_A(星期一)$=1$；μ_A(星期二)$=1$；μ_A(星期三)$=1$；μ_A(星期四)$=0$；μ_A(星期五)$=0$；μ_A(星期六)$=0$；μ_A(星期日)$=0$。

实际上，在好天气、坏天气之间是有差异的，利用模糊集合的概念可以选取 $[0,1]$ 之间的数，对天气情况加以细化。这时相对于好天气的隶属度可写成：μ_A(星期一)$=0.9$；μ_A(星期二)$=0.8$；μ_A(星期三)$=0.7$；μ_A(星期四)$=0.4$；μ_A(星期五)$=0.3$；μ_A(星期六)$=0.2$；μ_A(星期日)$=0.1$。

模糊集合表示法

普通集合的表示法有枚举法，即枚举所有元素，适用于元素个数有限的集合；还有描述法，也就是给出元素的属性或判定性质。这两种方法都是集合的元素表达法。除此，还可以使用隶属函数表达，这是一种对外延的表达。

对于模糊集合来说，存在多种表示方法，其原则是要反映出每个元素及其隶属度。设论域 $U=\{u_1,u_2,\cdots,u_n\}$，A 是其上的模糊子集。常见的表示方法有下述几种形式。

1. 扎德(Zadeh)表示法

$$A=\frac{\mu_A(u_1)}{u_1}+\frac{\mu_A(u_2)}{u_2}+\cdots+\frac{\mu_A(u_n)}{u_n}=\sum_{i=1}^{n}\frac{\mu_A(u_i)}{u_i} \tag{1.4}$$

其中符号“+”与“$\sum$”都不再是加号与求和号，而是衔接符号，而 $\mu_A(u_i)/u_i$ 也不表示分数，只是表示 u_i 对模糊集合 A 的隶属度。

对于任何论域 U，特别是 U 为无限集，扎德还采用“积分符号表示法”，即将 U 上的模糊子集统一地用

$$A=\int_U \frac{\mu_A(u)}{u} \tag{1.5}$$

其中$\int$并不表示积分，只是表示模糊集合的一种符号。

前面的模糊集合事例可写成

$$A=\frac{0.9}{星期一}+\frac{0.8}{星期二}+\frac{0.7}{星期三}+\frac{0.4}{星期四}+\frac{0.3}{星期五}+\frac{0.2}{星期六}+\frac{0.1}{星期日}$$

其中“分母”表示论域中的元素，而“分子”是相应元素的隶属度。

2.“向量”表示法

将论域 U 上的隶属度按顺序写成向量形式，比如前面的模糊集合可写成

$$A=\{0.9,0.8,0.7,0.4,0.3,0.2,0.1\}$$

对于一般的集合 A，可以表示为

$$A=\{\mu_1,\mu_2,\cdots,\mu_n\} \tag{1.6}$$

其中，$\mu_i[0,1]$，$i=1,2,\cdots,n$。μ_i 表示第 i 个元素对集合的隶属度。

3.“序偶”表示法

将元素与其对应的隶属度一起构成有序对的形式，前面的模糊集合事例可写成

$$A=\{(0.9,星期一),(0.8,星期二),(0.7,星期三),(0.4,星期四),(0.3,星期五),(0.2,星期六),(0.1,星期日)\}$$

对于一般情况，有

$$A=\{(\mu_1,u_1),(\mu_2,u_2),\cdots,(\mu_n,u_n)\} \tag{1.7}$$

4.“单点”表示法(称为单点)

$$A=\{\mu_1/u_1,\mu_2/u_2,\cdots,\mu_n/u_n\} \tag{1.8}$$

5.具体给出隶属函数的解析式

当论域为实数集上的一区间时，此法显得简单而方便。

【例1.2】 扎德给出年龄论域 $U=[0,100]$ 上的老年(Old)及青年(Young)两个模糊子集的隶属函数如下

$$\mu_O(u)=\begin{cases}0 & 当\ 0\leqslant u\leqslant 50\\ \left[1+\left(\dfrac{u-50}{5}\right)^{-2}\right]^{-1} & 当\ 50<u\leqslant 100\end{cases}$$

$$\mu_Y(u)=\begin{cases}1 & 当\ 0\leqslant u\leqslant 25\\ \left[1+\left(\dfrac{u-25}{5}\right)^{2}\right]^{-1} & 当\ 25<u\leqslant 100\end{cases}$$

其隶属函数曲线，如图1.1所示。

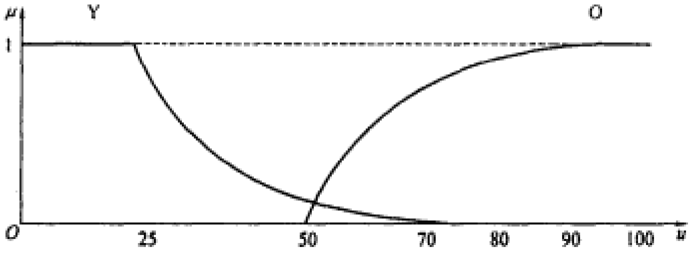


图 1.1 老年及青年的隶属函数

1.3 模糊集合运算

普通集合的运算可以利用特征函数来刻画及描述。由于隶属函数是特征函数的推广，所以模糊集合的运算也可以利用隶属函数描述。

【定义 1.2】 模糊集合运算

设 A,B,C,D 表示论域 U 上的模糊子集，可以定义出下述运算：

(1) 若对于任意的 $x\in U$，有 $\mu_A(x)\leqslant\mu_B(x)$，则称 B 包含 A，记为 $A\subseteq B$，如图 1.2 所示。

(2) 若 $A\subseteq B$，且 $B\subseteq A$，则称 A 与 B 相等，记为 $A=B$，如图 1.3 所示。

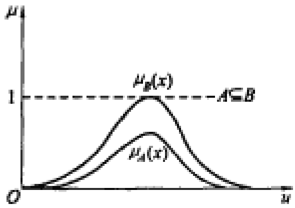


图 1.2

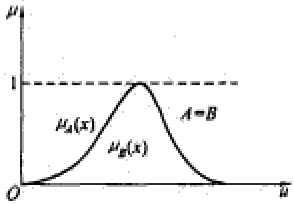


图 1.3

(3) 若对于任意的 $x\in U$，有

$$\mu_C(x)=\max\{\mu_A(x),\mu_B(x)\}=\mu_A(x)\vee\mu_B(x) \tag{1.9}$$

则称 C 为 A 与 B 的并集，记为 $C=A\cup B$，符号 $\vee$ 表示“取大(max)”运算，如图 1.4 所示。

(4) 若对于任意的 $x\in U$，有

$$\mu_D(x)=\min\{\mu_A(x),\mu_B(x)\}=\mu_A(x)\wedge\mu_B(x) \tag{1.10}$$

则称 D 为 A 与 B 的交集，记为 $D=A\cap B$，符号 $\wedge$ 表示“取小(min)”运算，如

图 1.5 所示。

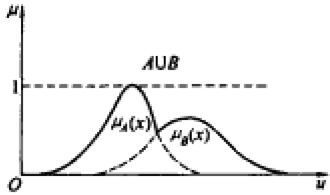


图 1.4

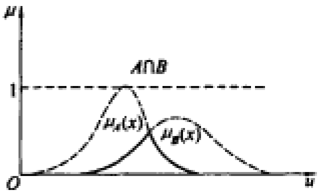


图 1.5

(5) 若对于任意的 $x \in U$,有

$$\mu_{A^c}(x) = 1 - \mu_A(x) \tag{1.11}$$

则称 A^c 为 A 的余集(或者补集),如图 1.6 所示。

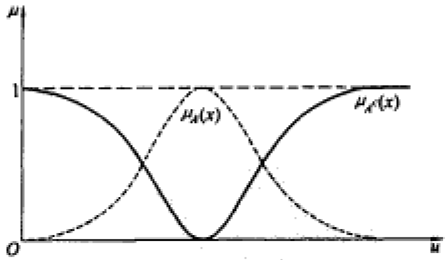


图 1.6

【例1.3】 设 A,B 表示两国人民对宗教的隶属度,其具体表达式可利用式(1.4)表示为

$$A = \frac{0.3}{佛教} + \frac{0.2}{基督教} + \frac{0.1}{伊斯兰教} + \frac{0.3}{其他教} + \frac{0.1}{未信教}$$

$$B = \frac{0.2}{佛教} + \frac{0.4}{基督教} + \frac{0.2}{伊斯兰教} + \frac{0.1}{其他教} + \frac{0.2}{未信教}$$

则根据定义 1.2 可得下述关系

$$A^c = \frac{0.7}{佛教} + \frac{0.8}{基督教} + \frac{0.9}{伊斯兰教} + \frac{0.7}{其他教} + \frac{0.9}{未信教}$$

$$A \cap B = \frac{0.2}{佛教} + \frac{0.2}{基督教} + \frac{0.1}{伊斯兰教} + \frac{0.1}{其他教} + \frac{0.1}{未信教}$$

$$A \cup B = \frac{0.3}{佛教} + \frac{0.4}{基督教} + \frac{0.2}{伊斯兰教} + \frac{0.3}{其他教} + \frac{0.2}{未信教}$$

为了简单起见,这里仅仅考察有限论域情况。至于无限论域的情况,可以作

出类似的讨论。

【定义 1.3】 模糊集合内积与外积

设 A,B 表示论域 U 上的模糊子集,把 A 与 B 的内积定义为

$$\bigvee_{x\in U}(\mu_A(x)\wedge\mu_B(x)) \tag{1.12}$$

记为 $A\circledcirc B$;把 A 与 B 的外积定义为

$$\bigwedge_{x\in U}(\mu_A(x)\vee\mu_B(x)) \tag{1.13}$$

记为 $A\otimes B$。

特别地,当 $\boldsymbol{A}=(a_1,a_2,\cdots,a_n)$ 与 $\boldsymbol{B}=(b_1,b_2,\cdots,b_n)$ 都是模糊向量时,$\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的内积、外积分别为

$$\boldsymbol{A}\circledcirc\boldsymbol{B}=\bigvee_{i=1}^{n}(a_i\wedge b_i),\quad \boldsymbol{A}\otimes\boldsymbol{B}=\bigwedge_{i=1}^{n}(a_i\vee b_i) \tag{1.14}$$

【性质 1.1】 $A\circledcirc B=1-A^c\circledcirc B^c,A\otimes B=1-A^c\otimes B^c$。

【性质 1.2】 $A\circledcirc A^c\leqslant\frac{1}{2},A\otimes A^c\geqslant\frac{1}{2}$。

【性质 1.3】 设 $A=(a_1,a_2,\cdots,a_n)$,令

$$\overline{A}=\bigvee_{i=1}^{n}a_i,\underline{A}=\bigwedge_{i=1}^{n}a_i$$

称 $\overline{A}$ 与 $\underline{A}$ 分别为 A 的高与低,则

$$A\circledcirc B\leqslant\overline{A}\wedge\overline{B},\quad A\otimes B\geqslant\underline{A}\wedge\underline{B}$$

【性质 1.4】 若 $A\subseteq B$,则

$$A\circledcirc B=\overline{A},A\otimes B=\underline{B}$$

模糊集合之间的接近程度,可以利用一种类似于距离的贴近度来描述。下面给出贴近度的公理化定义。

【定义 1.4】 模糊集合贴近度

设有 $\Phi(U)$ 上的二元函数

$$\tau:\Phi(U)\times\Phi(U)\to[0,1],(A,B)\mapsto\tau(A,B) \tag{1.15}$$

若 $\tau(A,B)$ 满足:

(1)$\tau(A,A)=0,\tau(\varnothing,U)=0$;

(2)$\tau(A,B)=\tau(B,A)$;

(3)$A\subseteq B\subseteq C\Rightarrow\tau(A,C)\leqslant\tau(A,B)\wedge\tau(B,C)$。

则称 τ 是 $F(U)$ 上的贴近度函数,称 $\tau(A,B)$ 为 A 与 B 的贴近度。

若把上述定义中的(1)换为(4),即

(4)$\tau(A,B)=1\Leftrightarrow A=B$,且 $\tau(\varnothing,U)=0$,则称 $\tau(A,B)$ 为 A 与 B 的严格贴近度。

通常，贴近度真实地反映出模糊集合的接近程度。但是，满足贴近度定义的函数并不是唯一的。一般地，常用的贴近度有下述几种：

(1) 利用距离表示贴近度

$$\tau(A,A) = 1 - c[d_p(A,B)]^a \tag{1.16}$$

其中 c 与 a 是两个可供选择的参数，而 $d_p(A,B)$ 定义如下

$$d_p(A,B) = \left(\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)|^p\right)^{\frac{1}{p}} \quad p \geqslant 1 \tag{1.17}$$

称为 A 与 B 的闵可夫斯基(Minkowski)距离。当 $p = 1$ 时，式(1.17)变成为

$$d_p(A,B) = \sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)| \tag{1.18}$$

式(1.18)称为 A 与 B 的汉明(Hamming)距离。当 $p = 2$ 时，式(1.17)变成为

$$d_p(A,B) = \left(\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)|^2\right)^{\frac{1}{2}} \tag{1.19}$$

式(1.19)称为 A 与 B 的欧几里得(Euclid)距离。

(2) 利用取大、取小表示贴近度

$$\tau(A,A) = \frac{\sum_{i=1}^{n}(\mu_A(x_i) \wedge \mu_B(x_i))}{\sum_{i=1}^{n}(\mu_A(x_i) \vee \mu_B(x_i))} \tag{1.20}$$

例如，对于前面的例1.3来说

$$d_1(A,B) = \frac{1}{5}\sum_{i=1}^{5} 1_{\{x_i \mid |\mu_A(x_i)-\mu_B(x_i)|>0\}}(x_i) = 1$$

$$d_2(A,B) = \frac{1}{\sqrt{5}}\sqrt{0.1^2 + 0.1^2 + 0.1^2 + 0.4^2 + 0.1^2} = 0.126$$

值得注意的是，d_1 通常用于属性的二元逻辑测度。

再比如，假定某个投资者对 X,Y,Z 房屋的评估为满意，不满意两种，即

	地　点	交　通	屋　况	环　境	价　格
X	满意	不满意	满意	满意	满意
Y	满意	满意	不满意	不满意	满意
Z	不满意	不满意	满意	满意	满意

则 $d_1(X,Y) = 0.6, d_1(X,Z) = 0.2, d_1(Y,Z) = 0.8$。

1.4 模糊矩阵

在现实世界中，事物之间的关系错综复杂，有时会出现多种相互关联的模糊现象，这就需要使用模糊矩阵来刻画。下面给出模糊矩阵的定义及运算。

【定义 1.5】 模糊矩阵

矩阵 $A=[a_{ij}]_{m\times n}$ 称为模糊矩阵，如果对任意的 $i\leqslant m$ 及 $j\leqslant n$，都有 $a_{ij}\in[0,1]$。如果 $a_{ij}\in\{0,1\}$，也就是元素只取 0 与 1 两个值的矩阵，那么称 A 为布尔矩阵。

【定义 1.6】 模糊矩阵运算

对于任意矩阵 $A=(a_{ij})_{m\times n}$ 与 $B=(b_{ij})_{m\times n}$，$A+B=(a_{ij}\vee b_{ij})$ 称为 A 与 B 的和；$A\times B=(a_{ij}\wedge b_{ij})$ 称为 A 与 B 的叉积，其中符号 $a\vee b$ 表示在 a 与 b 中取较大的，而符号 $a\wedge b$ 表示在 a 与 b 中取较小的。若 $a_{ij}\leqslant b_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$，则称 A 小于等于 B，记为 $A\leqslant B$。

由定义 1.6 知，两个模糊矩阵只有在行列分别相等的情况下才能够相加、叉乘以及比较大小。

通常，对于任意 m 行 n 列的模糊矩阵 A,B,C 来说，具有下述一些性质：

(1) 交换律：$A+B=B+A,A\times B=B\times A$；

(2) 结合律：$(A+B)+C=A+(B+C),(A\times B)\times C=A\times(B\times C)$；

(3) 分配律：$(A+B)\times C=(A\times C)+(B\times C),(A\times B)+C=(A+C)\times(B+C)$；

(4) 幂等律：$A+A=A,A\times A=A$；

(5) 吸收律：$(A+B)\times B=B,(A\times B)+B=B$；

(6) 对于任意 m 行 n 列的模糊矩阵 A，都有

$$0\leqslant A\leqslant E,0+A=A,0\times A=0,E+A=A,E\times A=A$$

$$A\leqslant B\Leftrightarrow A+B=B\Leftrightarrow A\times B=A$$

其中 $$0=\begin{bmatrix}0&0&\cdots&0\\0&0&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&0\end{bmatrix},\quad E=\begin{bmatrix}1&1&\cdots&1\\1&1&\cdots&1\\\vdots&\vdots&&\vdots\\1&1&\cdots&1\end{bmatrix}$$

(7) 若 $A_1\leqslant B_1,A_2\leqslant B_2$，则

$$A_1+A_2\leqslant B_1+B_2,A_1\times A_2\leqslant B_1\times B_2$$

【定义 1.7】　模糊矩阵乘积运算

设 m 行 n 列的模糊矩阵 A 与 n 行 p 列的模糊矩阵 B 分别为

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},\quad B=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}$$

它们的乘积 C 为 m 行 p 列的模糊矩阵，亦即

$$C=AB=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{bmatrix}$$

其中

$$c_{ij}=(a_{i1}\wedge b_{i1})\vee(a_{i2}\wedge b_{i2})\vee\cdots\vee(a_{in}\wedge b_{in})=\bigvee_{k=1}^{n}(a_{ik}\vee b_{ik}) \tag{1.21}$$

【例 1.4】

$$\begin{bmatrix} 0.8 & 0.2 & 0.1 & 0.5 \\ 0.6 & 0.9 & 0.7 & 0.4 \\ 0.1 & 0.3 & 0.5 & 0.8 \end{bmatrix}\begin{bmatrix} 0.6 & 0.3 \\ 0.5 & 0.9 \\ 0.4 & 0.1 \\ 0.7 & 0.4 \end{bmatrix}=\begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.9 \\ 0.7 & 0.4 \end{bmatrix}$$

模糊矩阵的乘法具有下述几个性质：

(1)$(AB)C=A(BC)$，$A^mA^n=A^{m+n}$，$(A^m)^n=A^{mn}$；

(2)$(A+B)C=AC+BC$，$C(A+B)=CA+CB$；

(3) 若 $A\leqslant B$，则 $AC\leqslant BC$，$CA\leqslant CB$，$A^n\leqslant B^n$；

(4) 若 $A\leqslant B$，$C\leqslant D$，则 $AC\leqslant BD$；

(5)$A0=A0=0$，$IA=AI=I$。

其中 I 与普通矩阵中的单位阵相同。

值得注意的是，$AB\neq BA$，$(A\times B)C\neq AC\times BC$，$C(A\times B)\neq CA\times CB$，一般地有 $(A\times B)C\leqslant AC\times BC$，$C(A\times B)\leqslant CA\times CB$。

1.5 截集与支集

模糊集合是普通集合的推广，普通集合是模糊集合的特殊形式。在一定条件下，两者可以互相转化。利用水平截集可以将模糊集合转换成普通集合。支集在模糊集合的实际应用中起着十分重要的作用。

【定义1.8】 设 A 为论域 U 上的模糊集合，对任意 $\lambda \in [0,1]$，称

$$A_\lambda = \{u \mid u \in U, \mu_A(u) \geqslant \lambda\} \tag{1.22}$$

为 A 的λ 截集（λ - cut，或称为 λ 水平截集），而 λ 称为置信水平。若把定义式中不等号"$\geqslant$"换成"$>$"，则称这样的 A_λ 为 A 的强截集。

由定义知，通过对 A 的隶属度确定一个限定值$\lambda(0 \leqslant \lambda \leqslant 1)$，然后再把隶属度 $\mu_A(u) \geqslant \lambda$ 的元素挑出来，就得到普通集合。

λ 水平截集是普通集合，其特征函数为

$$I_{A\mu}(u) = \begin{cases} 1 & 当\ \mu_A(u) \geqslant \lambda \\ 0 & 当\ \mu_A(u) < \lambda \end{cases}$$

其中 $I_{A\mu}(u)$ 与 $\mu_A(u)$ 关系，如图 1.7 所示，图中粗曲线表示 A_μ 的特征函数。

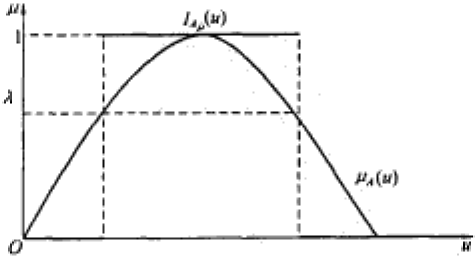


图 1.7 A_μ 的特征函数曲线

【定义1.9】 设 A 为论域 U 上的模糊集合，把 $\mu_A(u) = 1$ 的所有 u 组成的集合，即

$$A_1 = \text{Ker } A = \{u \mid \mu_A(u) = 1\} \tag{1.23}$$

称为 A 的核集，记为 A_1 或 Ker A。

为了研究方便，把

$$A_\alpha = \{u \mid \mu_A(u) = \alpha\} \tag{1.24}$$

称为水平集合（α - level）。

【定义1.10】 设 A 为论域 U 上的模糊集合，把 $\mu_A(u) > 0$ 的所有 u 组成

的集合，即

$$\text{Supp } A = \{u \mid \mu_A(u) > 0\} \tag{1.25}$$

称为 A 的支集(又称为支撑集)，记为 Supp A。

若 Ker $A \neq \varnothing$，则称 A 为正规模糊集合。否则，称 A 为非正规模糊集合。随着 λ 由 1 变到 0，A_λ 从 $A_\lambda = 1$ 出发不断扩大，收进来的元素愈来愈多，达到 Supp $A = \{u \mid \mu_A(u) > 0\}$，是隶属度大于 0 的元素的最大集合。把 Supp A − Ker A 称为 A 的边界。A_λ，Ker A 以及 Supp A 之间的关系，如图 1.8 所示。

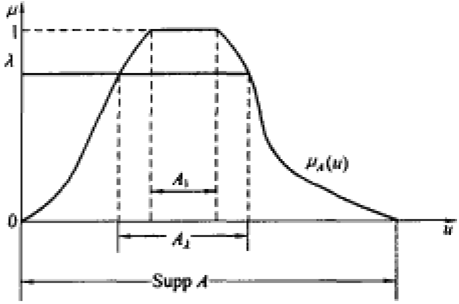


图 1.8 A_λ，A_1 以及 Supp A 关系图

下面介绍分解定理与扩展原理。分解定理和扩展原理是模糊数学的理论支柱。分解定理是联系模糊数学和普通数学的纽带，而扩展原理是把普通的数学方法扩展到模糊数学中的有力工具。

【定理 1.1】 分解定理

假设 $\tilde{A}$ 为定义在论域 X 上的模糊集，则有

$$\mu_{\tilde{A}}(x) = \max_{\alpha\in[0,1]} \min\{\alpha, \mu_{A_\alpha}(x)\} \tag{1.26}$$

其中，$\mu_{A_\alpha}(x)$ 是 A_α 的特征函数。

证明

$$\mu_{A_\alpha}(x) = \begin{cases} 1 & x \in A_\alpha (\text{即 } \mu_{\tilde{A}}(x) \geqslant \alpha) \\ 0 & x \notin A_\alpha (\text{即 } \mu_{\tilde{A}}(x) < \alpha) \end{cases}$$

故而

$$\max_{\alpha\in[0,1]} \min\{\alpha, \mu_{A_\alpha}(x)\} = \max_{\alpha\leqslant\mu_{\tilde{A}}(x)} \alpha = \mu_{\tilde{A}}(x)$$

从而定理得证。

如果引入逻辑符号“∨”和“∧”，式(1.26)也可以等价地表示为

$$\mu_{\tilde{A}}(x) = \bigvee_{\alpha\in[0,1]} \{\alpha \wedge \mu_{A_\alpha}(x)\} \tag{1.27}$$

由此可见，知道了模糊集合 $\tilde{A}$，就可以获得它的 α 截集；反之，知道了 $\tilde{A}$ 的全体 α 截集，也可以合成 $\tilde{A}$。

分解定理的内容，可由图 1.9 给出直观说明。图中画出不同水平的 $\mu_{\lambda A_\lambda}(x)$ 的图形。当 λ 取遍[0,1]闭区间所有值时，$\bigcup \lambda A_\lambda$ 按模糊子集求并运算法则，也就是取各 $\lambda \in [0,1]$ 点隶属函数的最大值，再连成一条曲线，这自然与 $\mu_{\tilde{A}}(x)$ 曲线重合，分解定理就说明了这个道理。

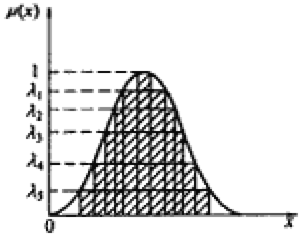


图 1.9　分解定理的图示

【定义 1.11】　假设 $A \subseteq X$ 是论域 X 中的普通集合，$\alpha \in [0,1]$ 是一个实数，则符号 $\alpha \cdot A$ 表示由 α 和 A 所确定的一个模糊集，其隶属函数是

$$\mu_{\alpha \cdot A}(x) = \begin{cases} \alpha & x \in A \\ 0 & x \notin A \end{cases} \tag{1.28}$$

这个模糊集称为 α 与 A 的"积"。

用上面的符号，分解定理可方便地叙述为：设 $\tilde{A}$ 为论域 X 上的模糊集，则

$$\tilde{A} = \bigcup_{\alpha \in [0,1]} (\alpha \cdot A_\alpha) \tag{1.29}$$

【定义 1.12】　集合在映射下的像

设有映射 $f: X \to Y$，且 $A \subseteq X$，那么定义

$$f(A) = \{y \mid \exists x \in A, y = f(x)\} \tag{1.30}$$

为集合 A 在 f 下的像。

式(1.30)也可以用特征函数表示为

$$\mu_{f(A)}(y) = \begin{cases} \bigvee\limits_{x \in f^{-1}(y)} \mu_A(x) & f^{-1}(y) \neq \varnothing \\ 0 & f^{-1}(y) = \varnothing \end{cases} \tag{1.31}$$

集合在映射下的像是非模糊的数学概念，为了能够处理模糊量值的问题，需要对这一概念进行扩展。给定映射 $f: X \to Y$，以及 X 中的模糊集合 $\tilde{A}$，应如何确定 $\tilde{A}$ 在 f 下的像？扩展原理对此作了回答。扩展原理在模糊集理论中有着重要

的作用。它的实质是一个定义。

【定义1.13】 模糊集合在映射下的像(扩展原理)

设有映射 $f: X \to Y$,且 $\tilde{A}$ 是 X 中的模糊集合,那么记 $\tilde{A}$ 在 f 下的像为 $f(\tilde{A})$,它是 Y 中的模糊集合,且具有如下的隶属函数

$$\mu_{f(\tilde{A})}(y) = \begin{cases} \bigvee\limits_{x \in f^{-1}(y)} \{\mu_{\tilde{A}}(x)\} & f^{-1}(y) \neq \varnothing \\ 0 & f^{-1}(y) = \varnothing \end{cases} \tag{1.32}$$

当 f 为一一映射时,式(1.32)可以简化为

$$\mu_{f(\tilde{A})}(y) = \begin{cases} \mu_{\tilde{A}}(x) & f^{-1}(y) \neq \varnothing \\ 0 & f^{-1}(y) = \varnothing \end{cases} \tag{1.33}$$

考虑更一般的情况,设有一个 n 元函数 f(它是从笛卡儿乘积 $X_1 \times X_2 \times \cdots \times X_n$ 到 Y 的一个映射)和 $X_1, X_2, \cdots, X_n$ 中的 n 个模糊集合 $\tilde{A}_1, \tilde{A}_2, \cdots, \tilde{A}_n$。我们来给出这种情况下扩展原理的表达式。

【定义1.14】 设 $X = X_1 \times X_2 \times \cdots \times X_n$,$\tilde{A}_1, \tilde{A}_2, \cdots, \tilde{A}_n$ 相应地是 $X_1, X_2, \cdots, X_n$ 中的模糊集合。$\tilde{A}_1, \tilde{A}_2, \cdots, \tilde{A}_n$ 的笛卡儿乘积集记为 $\tilde{A}_1 \times \tilde{A}_2 \times \cdots \times \tilde{A}_n$,定义为 X 中的模糊集合,它的隶属函数为

$$\mu_{\tilde{A}_1 \times \tilde{A}_2 \times \cdots \times \tilde{A}_n}(x_1, x_2, \cdots, x_n) = \wedge \{\mu_{\tilde{A}_1}(x), \mu_{\tilde{A}_2}(x), \cdots, \mu_{\tilde{A}_n}(x)\} \tag{1.34}$$

【定义1.15】 扩展原理的一般表示式

设有映射 $f: X_1 \times X_2 \times \cdots \times X_n \to Y$,它给出 $y = f(x_1, x_2, \cdots, x_n)$,且 $\tilde{A}_1, \tilde{A}_2, \cdots, \tilde{A}_n$ 相应地是 $X_1, X_2, \cdots, X_n$ 中的模糊集合,那么扩展原理断言,由 $\tilde{A}_1 \times \tilde{A}_2 \times \cdots \times \tilde{A}_n$ 通过 f 导出的 Y 中的模糊集合 $\tilde{B}$ 具有隶属函数

$$\mu_{\tilde{B}}(y) = \begin{cases} \bigvee\limits_{f(x_1, x_2, \cdots, x_n) = y} \{\mu_{\tilde{A}_1 \times \tilde{A}_2 \times \cdots \times \tilde{A}_n}(x_1, x_2, \cdots, x_n)\} & f^{-1}(y) \neq \varnothing \\ 0 & f^{-1}(y) = \varnothing \end{cases} \tag{1.35}$$

根据式(1.34),上式又可写为

$$\mu_{\tilde{B}}(y) = \begin{cases} \bigvee\limits_{f(x_1, x_2, \cdots, x_n) = y} \{\mu_{\tilde{A}_1}(x_1) \wedge \mu_{\tilde{A}_2}(x_2) \wedge \cdots \wedge \mu_{\tilde{A}_n}(x_n)\} & f^{-1}(y) \neq \varnothing \\ 0 & f^{-1}(y) = \varnothing \end{cases} \tag{1.36}$$

当 $n = 1$ 时,式(1.36)简化为式(1.32)。

【例1.5】 设 $A = \frac{0.4}{u_1} + \frac{0.8}{u_2} + \frac{0.7}{u_3} + \frac{0.5}{u_4} + \frac{0}{u_5} + \frac{0.9}{u_6} + \frac{1}{u_7}$,有:

$\lambda = 1$ 时，$A_1 = \{u_7\}$；

$\lambda = 0.9$ 时，$A_{0.9} = \{u_6, u_7\}$；

$\lambda = 0.7$ 时，$A_{0.7} = \{u_3, u_6, u_7\}$；

$\lambda = 0.6$ 时，$A_{0.6} = \{u_3, u_6, u_7\}$；

$\lambda = 0.5$ 时，$A_{0.5} = \{u_3, u_4, u_6, u_7\}$；

$\lambda = 0$ 时，$A_0 = \{u_5\}$。

【定义 1.16】 对于任意的 $\lambda \in [0,1]$，如果模糊矩阵 $\boldsymbol{A} = [a_{ij}]_{m\times n}$ 的元素 a_{ij} 满足

$$a_{ij}^{\lambda} = \begin{cases} 1 & a_{ij} \geqslant \lambda \\ 0 & a_{ij} < \lambda \end{cases}$$

那么称由 a_{ij}^{λ} 所构成的矩阵为 $\boldsymbol{A} = [a_{ij}]_{m\times n}$ 的 λ 截矩阵，显然，A_λ 的元素要么是 1，要么是 0，因此是一个布尔矩阵。

【例 1.6】 设模糊矩阵为

$$\boldsymbol{A} = \begin{pmatrix} 0.8 & 0.9 & 0.6 & 0 \\ 0.1 & 1 & 0.2 & 0.4 \\ 0.5 & 0.7 & 0.8 & 0.3 \end{pmatrix}$$

则当 $\lambda = 0.5$ 时，其 λ 截矩阵为

$$\boldsymbol{A}_{0.5} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

如同以后将看到的，使用保持普通集的凸性所具有的许多性质的方法能够容易地将凸性的许多性质扩展到模糊集上。在模式分类、优化和有关问题等的应用中，凸概念特别有用。

【定义 1.17】 设 A 为论域 $U = \mathbf{R}$ 上的模糊集合，模糊集 A 是凸的，当且仅当集合

$$A_\alpha = \{x \mid f_A(x) \geqslant \alpha\} \tag{1.37}$$

对于所有的 $\alpha \in (0,1]$ 而言是凸的。一个可选择的更为直接的定义如下：A 是凸的，当且仅当

$$f_A[\lambda x_1 + (1-\lambda)x_2] \geqslant \min[f_A(x_1), f_A(x_2)] \tag{1.38}$$

对所有 $x_1, x_2 \in X$ 以及对所有的 $\lambda \in [0,1]$ 成立。注意，这个定义并不意味着 $f_A(x)$ 必须是 x 的凸函数，如图 1.10 中 $n = 1$ 的情况。

注意到，上述二者定义之间的等价性如下：若 A 在第一个定义之下是凸的，

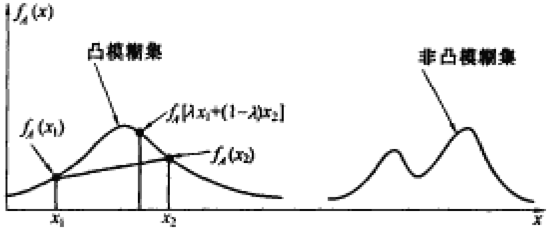


图 1.10 在 **R** 上的凸和非凸模糊集

且有

$$\alpha = f_A(x_1) \leqslant f_A(x_2)$$

则由 A_α 的凸性定义得出 $x_2 \in A_\alpha$ 和 $\lambda_1 x_1 + (1-\lambda)x_2 \in A_\alpha$。因此

$$f_A[\lambda x_1 + (1-\lambda)x_2] \geqslant \alpha = f_A(x_1) = \min[f_A(x_1), f_A(x_2)]$$

相反的，若 A 在第二个定义之下是凸的，而且 $\alpha = f_A(x_1)$，则可以把 A_α 看做所有满足 $f_A(x_2) \geqslant f_A(x_1)$ 的点之集合。由于式(1.37) 具有形式 $\lambda x_1 + (1-\lambda)x_2(0 \leqslant \lambda \leqslant 1)$ 的每个点也在 A_α，因此 A_α 是凸集。

凸模糊集具有以下两条性质：

(1) 凸模糊集的截集必是区间；截集均为区间的模糊集必为凸模糊集。

(2) $\tilde{A}$，$\tilde{B}$ 是凸模糊集，则 $\tilde{A} \cap \tilde{B}$ 也是凸模糊集。

【定义 1.18】 设 A 为实数论域 **R** 上的一个模糊集，若 $\forall \lambda \in [0,1]$，$A_\lambda$ 是 **R** 中有限闭区间，则称 A 为 **R** 上的一个模糊数；若 A 为模糊数，且 $A_1 = \text{Ker } A = \{a\}$，则称 A 是关于 a 的严格模糊数。

模糊数是定义在实数域上的一类特殊的模糊集合，它在模糊数学和模糊优化的理论及应用研究中起着十分重要的作用。

由上述定义看出，模糊数是实数轴上的一个闭的区间数。在实际的工程与社会经济系统中，许多因素不能用准确的数量来刻画，但是人们却知道这些量的范围及属于该范围的程度。模糊数为这些因素的描述提供了一个有效的数学工具。实际上，除了上面的定义之外，模糊数还可以定义成实数域上的具有凸的分段连续的隶属函数的模糊集。

通常，最常用的模糊数是三角形模糊数与梯形模糊数，它们都是依据其隶属函数的几何形状而命名的。另外，还有其他形式的模糊数，比如 LR 模糊数(可参看 D. Dubuis and H. Prade, Fuzzy Sets and Systems: Theory and Applications, Academic Press)。

第 2 章　　隶属函数及软计算方法

隶属函数通常可分为离散型(Discrete Type)与连续型(Continuous Type)两类。离散型隶属函数是直接对有限模糊集合内每个元素给定隶属度，并以"向量"形式表现出来，而连续型隶属函数则有几种常用的函数形式，诸如 S 型函数、Z 型函数、π 型函数、三角形函数、梯形函数、高斯(钟形)函数来描述模糊集合。函数定义的表现，可以是无限模糊集合的元素及其隶属度之间的关系，也可以是有限模糊集合的元素及其隶属度之间的关系。

2.1　隶属函数

过去我们利用普通集合定义具有模糊性质的语言变量(Linguistic Variable，又称语言变项，可参看 2.2 节)时，经常会造成许多不合理的现象。例如，最近相当热门的"景气"一词，当我们考虑景气指标在 0 到 100 的范围时，若定义 30 到 50 为景气好，则根据普通集合的定义，可绘出景气的特征函数，如图 2.1 所示。

在图 2.2 中显示，当某月景气指标介于 30 到 50 之间，则属于景气好，其特征值为 1；反之，便不属于景气好，特征值为 0。但是，若我们假定有 A，B，C 三个月，景气指标各为 29，31，49，其中 B，C 两月指标值相差 18，且都属于景气好；但 A，B 两月指标值虽然只差 2，但 A 月却不属于景气好，这是相当不合理的。故并不是全部都合理，须看其特性决定。

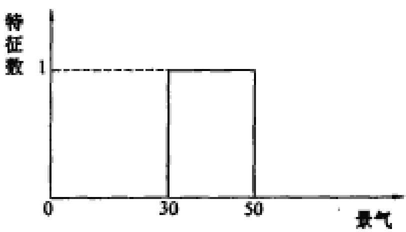


图 2.1　普通集合：景气的特征函数

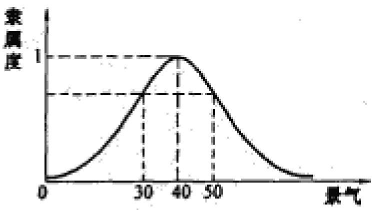


图 2.2　模糊集合：景气的隶属函数

对于这种普通集合的二分法与人类思维模式格格不入的问题，利用隶属函数则能获得较为合理的答案。如果某个人认为 40 绝对属于景气好，那么其隶属函数值自然属于 1，而 39 或 41 的隶属函数值则大约等于 0.9，此表示 39 或 41 属于景气好的程度有 0.9 之多。根据隶属函数的定义，我们可绘出模糊集合景气指标的隶属函数。与普通集合的特征函数相比，隶属函数似乎是将特征函数平滑化了。不仅如此，隶属函数让每个景气指标都拥有一个介于 0 到 1 之间的值，来代表属于景气好的程度。与普通集合的特征函数相比，在描述模糊概念时，利用模糊集合的隶属函数来解释是更为恰当的。

隶属函数是模糊理论最基本的概念，它不仅可以描述模糊集合的性质，而且可以对模糊集合进行量化，并且利用精确的数学方法，来分析和处理人类模糊性的信息。然而，要建立一个足以表达模糊概念的隶属函数，并不是一件容易的事。其原因在于隶属函数仍旧脱离不了个人的主观意识，故没有通用的定理或公式。

目前，如何确定隶属函数还没有定法，具体问题具有特殊性。下面介绍几种经常使用的、具有代表性的方法。

模糊统计试验法

在概率论中，事件发生的概率可以通过大量的随机试验来求得。对随机试验的一个基本要求是，每次试验时事件的结果必须是确定的，假定结果有 A_1 和 A_2 两种可能，要么 A_1，要么 A_2，不能模棱两可，即不存在 A_1 与 A_2 的过渡状态。这次试验中，事件 A 发生的频率 = A 发生的次数/n。当 $n \to \infty$ 时，频率就是概率。

计算隶属度时，可采用类似于求概率的方法，就是模糊统计试验法，或简称为模糊统计法。

当我们进行试验时，会发现它呈现出一定的统计规律，亦即隶属频率的稳定性。

设 $A \in \Phi(U)$，$u_0 \in U$，现在的问题是，如何利用模糊统计实验来确定 u_0 对 A 的隶属度，即 $\mu_A(u_0)$。

如果我们作多次模糊统计试验，以较为确定的方式判定 U 中的哪些元素属于 A。显然，由于每次试验的条件不同(带有模糊性)，那么每次哪些元素被判定为属于 A 是不大明确的，它们构成 U 上的一个普通集合，记为 A^*，它通常是边界可变的、可移动的。当然，是否每一次 u_0 在 A^* 中很难说，即 $u_0 \in A^*$，也可

$u_0 \notin A^*$。

若作 n 次这样的试验，其中 m 次 $u_0 \in A^*$，我们称 m/n 为 u_0 对 A 的隶属频率 t，即

$$t = \frac{m}{n} = \frac{u_0 \in A^* \text{ 的次数}}{\text{总次数}} \tag{2.1}$$

随着 n 增大，隶属频率会出现稳定性，而稳定所在的那个数便定为 u_0 对 A 的隶属度 $\mu_A(u_0)$。对论域中每个元素不断运用上述方法，最终会获得论域 U 上的模糊子集 A 的隶属度。

模糊统计试验归纳起来，通常具有四个要素：

(1) 论域 U；

(2) 论域 U 中的一个元素 u_0；

(3) 论域 U 中的一个边界可变的普通集合 A^*，它体现出模糊概念外延的不确定型，亦即边界不明确；

(4) 试验条件，它体现出在模糊概念 A 影响下的划分过程，且制约着 A^* 的运动。

模糊试验的特点是，在各次试验中，u_0 是固定的，而 A^* 在随机运动。如果认为概率统计是考察"变动的点"是否落在"不动的圈内"，那么可以把模糊统计比喻为"变动的圈"是否落在"不动的点"。

【例 2.1】 本事例所引用的模糊统计试验源于我国学者张南论等人在武汉建材学院对模糊集合"青年人"所作的抽样试验。

设 $U = [0,1]$，A 为青年人，$u_0 = 27$ 岁。对 129 人作抽样调查，让各人给出"青年人"比较合适的年龄段。最后，整理出反映 27 岁属于"青年人"的隶属频率表，如表 2.1 所示。

表 2.1 隶属频率表

n	10	20	30	40	50	60	70	80	90	100	110	120	129
m	6	14	23	31	39	47	53	62	68	76	85	95	101
t	0.6	0.7	0.77	0.78	0.78	0.78	0.76	0.78	0.76	0.76	0.75	0.79	0.78

可见，隶属频率基本上是稳定在 0.78，故而得到 27 岁属于"青年人"的隶属度

$$\mu_A(27) = 0.78$$

按这种方法来求出各 $\mu_A(u)$，$u \in U$，可给出一条基本上以 18 ~ 25 岁为中心的正态型"青年人"隶属曲线。

以上讨论是试验结果对论域作二划分，即结果是属于 A 或 $\bar{A}$，从划分的意义上讲，有 $A \cup \bar{A} = U$。我们把这种模糊统计试验法称为二项模糊统计。类似地，有多项模糊统计，亦即 $A_1 \cup A_2 \cup \cdots \cup A_m = U$。若以前面所给论域 U，要同时统计出 u_0 属于子集"老年""中年""青年""少年" 的隶属度，就是四相模糊统计问题。

二元对比排序法

二元对比排序法是把事物两两相比，从而确定顺序，由此决定隶属函数的大致形状。

二元对比排序法有多种，比如：(1) 相对比较法；(2) 择优比较法；(3) 对比平均法。下面具体讲述相对比较法。

该方法是把论域中各元素，按某种特性，在两两元素之间进行对比，建立所谓二元比较级。设数对 $(f_y(x), f_x(y))$ 是 $[0,1] \times [0,1]$ 上的数偶，在二元比较级后，通过相似矩阵转化为总体的排序。

【例 2.2】　设由三个儿子组成的集合 $X = \{x_1, x_2, x_3\}$，其中 x_1 表示长子，x_2 表示次子，x_3 表示三子。试确定三个儿子与父亲的相似程序，即隶属度。

首先，先在两两元素之间进行二元比较，得

$$f_{x_1}(x_1) = 1, f_{x_2}(x_1) = 0.8, f_{x_3}(x_1) = 0.5$$

$$f_{x_1}(x_2) = 0.5, f_{x_2}(x_2) = 1, f_{x_3}(x_2) = 0.4$$

$$f_{x_1}(x_3) = 0.3, f_{x_2}(x_3) = 0.7, f_{x_3}(x_3) = 1$$

$(f_{x_2}(x_1), f_{x_1}(x_2)) = (0.8, 0.5)$ 是一个二元比较级，它的含义是把长子与次子相比较，如果长子与父亲的相似程度为 0.8，则次子与父亲的相似程度只有 0.5。同理可知其他二元比较级。$f_{x_1}(x_1) = f_{x_2}(x_2) = f_{x_3}(x_3) = 1$ 表示自己与自己比较与父亲相似的程度。因此，该问题的三个二元比较级是

$$(f_{x_2}(x_1), f_{x_1}(x_2)) = (0.8, 0.5)$$

$$(f_{x_3}(x_2), f_{x_2}(x_3)) = (0.4, 0.7)$$

$$(f_{x_3}(x_1), f_{x_1}(x_3)) = (0.5, 0.3)$$

然后，建立相似矩阵。令

$$f(x/y) = \frac{f_y(x)}{\max(f_x(y), f_y(x))}$$

显然有

$$f(x/y)=\begin{cases} f_y(x)/f_x(y) & \text{当 } f_y(x)\leqslant f_x(y) \\ 1 & \text{当 } f_y(x)>f_x(y) \end{cases}$$

此处 x,y 属于论域 X，以 $f(x/y)$ 为元素作成矩阵，且凡是 $f(x/x)$ 均为 1。这样构成的矩阵就叫做相似矩阵，上述问题的相似矩阵如下

$$S=\begin{matrix} & x_1 & x_2 & x_3 & \min \\ x_1 & 1 & 1 & 1 & 1 \\ x_2 & 5/8 & 1 & 4/7 & 4/7 \\ x_3 & 3/5 & 1 & 1 & 3/5 \end{matrix}$$

其次，排序规则：从 S 的每一行取最小值，按所得数从大到小排出优劣次序为 $1>\frac{3}{5}>\frac{4}{7}$，亦即 $1>0.6>0.57$。

其他方法

其他一些常见的方法还包括：评分法（专家评分直接给出隶属函数）；范例法（Zadeh，1972，依据有关的 μ_A 部分信息确定出 μ_A）；范形变形法（Bremermann，1976）；三分法（一种利用随机区间的思想来处理模糊性的试验模型）；优选法（Saaty，1974）；子集比较法（Fuug 和 Li，1974）；隐分析法（Kochen 和 Badre，1976）；滤波函数法（MacVicar-Whelam，1978）；模糊测度法（管野道夫）等。总之，要具体问题具体分析，在实践中不断增长经验。

若模糊集合定义在实数域上，则模糊集合的隶属函数就称为模糊分布。为了在处理这类问题时便于参考，下面给出一些经常使用的模糊分布。

类似于数理统计学中确定分布的方法，研究者可根据实际情况，选定一些带有参数的隶属函数表示某类型的模糊概念（通常论域为实数域），然后通过实验确定参数。这是应用研究者经常采用的方法。

中间型隶属函数

(1) 矩阵分布

$$\mu(x)=\begin{cases} 0 & x\leqslant a-b \\ 1 & a-b<x\leqslant a+b \\ 0 & a+b<x \end{cases}$$

参看图 2.3。

(2) 尖 Γ 分布

$$\mu(x)=\begin{cases} e^{k(x-a)} & x \leqslant a \\ e^{-k(x-a)} & x > a \end{cases}$$

参看图 2.4。

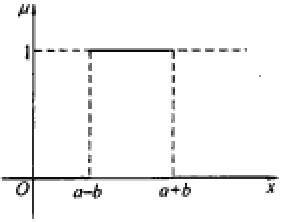


图 2.3　矩阵分布

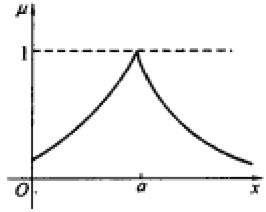


图 2.4　尖 Γ 分布

(3) 正态分布或高斯分布 $\mu(x)=e^{-(k-a)^2}$,其中 $k>0$。参看图 2.5。

(4) 柯西分布 $\mu(x)=\dfrac{1}{1+\alpha(x-a)^{\beta}}$,其中 $\alpha>0$,β 为偶数。参看图 2.6。

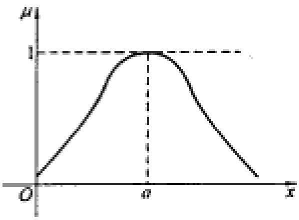


图 2.5　正态分布或高斯分布

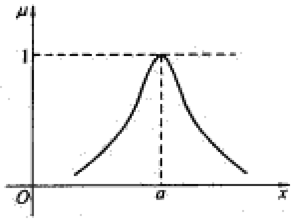


图 2.6　柯西分布

(5) 梯形分布

$$\mu(x)=\begin{cases} 0 & x \leqslant a-a_2 \\ \dfrac{a_2+x-a}{a_2-a_1} & a-a_2 < x \leqslant a-a_1 \\ 1 & a-a_1 \leqslant x < a+a_1 \\ \dfrac{a_2-x+a}{a_2-a_1} & a+a_1 \leqslant x < a+a_2 \\ 0 & a+a_2 \leqslant x \end{cases}$$

参看图 2.7。

(6) 岭形分布

$$\mu(x)=\begin{cases}0 & x\leqslant -a_2\\ \frac{1}{2}+\frac{1}{2}\sin\frac{\pi}{a_2-a_1}\left[x-\frac{a_2+a_1}{2}\right] & -a_2<x\leqslant -a_1\\ 1 & -a_1<x\leqslant a_1\\ \frac{1}{2}-\frac{1}{2}\sin\frac{\pi}{a_2-a_1}\left[x-\frac{a_2+a_1}{2}\right] & a_1<x\leqslant a_2\\ 0 & a_2<x\end{cases}$$

参看图 2.8。

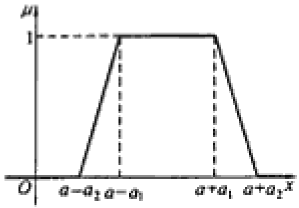


图 2.7 梯形分布

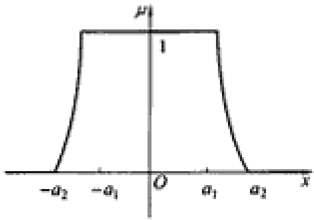


图 2.8 岭形分布

下面,再给出三种标准的隶属函数。

(7)S 型隶属函数

$$S(x;a,b)=\begin{cases}0 & x\leqslant a\\ 2\left[\frac{x-a}{b-a}\right]^2 & a<x\leqslant\frac{a+b}{2}\\ 1-2\left[\frac{x-b}{b-a}\right]^2 & \frac{a+b}{2}<x\leqslant b\\ 1 & b<x\end{cases}$$

参看图 2.9。

(8)Z 型隶属函数

$$Z(x;a,b)=1-S(x;a,b)$$

参看图 2.10。

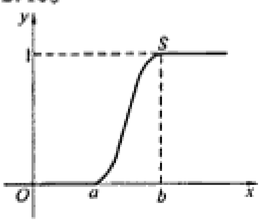


图 2.9 S 型隶属函数

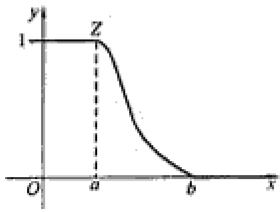


图 2.10 Z 型隶属函数

(9)π型隶属函数

$$\pi(x;a,b)=\begin{cases}S(x;b-a,b) & x\leqslant b\\ Z(x;b,a+b) & x>b\end{cases}$$

参看图2.11。

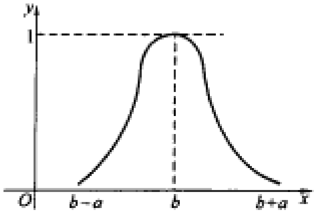


图2.11　π型隶属函数

至于偏大与偏小型隶属函数公式，参看附录1。

2.2　模糊数据与软计算

软计算(Soft Computing)与传统实数计算方法(也就是硬计算)不同的是，软计算是基于模糊数或模糊样本，包括区间数、多值数、语言变量等的数学运算。软计算的宗旨是，通过对不精确性、不确定性、部分真值以及近似表达的描述，使得问题变得易于处理、获得稳健性，减少求解费用，更好地与实际相符。

举一个事例，某大学生每周运动2～4次，而每次1～2小时，那么一周运动几小时?又比如，某人外出旅游7～8天，需携带若干胃肠药，而胃肠药标示成人每日3～4次，而每次5～7颗，请问他须携带多少颗。传统数学无法对此作出明确的计算，但是一般人均会作出大概的计算。这种事例就应属于软计算。

下面介绍本书中经常使用的三种形式的模糊数据(有时，也简称为模糊数)：区间数、模糊数以及三角模糊数。

区间数及其运算

设$\mathbf{R}$表示所有实数组成的集合，对于$a_1,a_2\in\mathbf{R}$，其中$a_1<a_2$形成一个区间。当不确定的x值满足$a_1\leqslant x\leqslant a_2$时，则$[a_1,a_2]$称为闭区间。$a_1,a_2$分别是区间的上界与下界。而$(a_1,a_2)$，$[a_1,a_2)$以及$(a_1,a_2]$分别称为有界开或闭区间。同时，称$(-\infty,a_2]$，$(-\infty,a_2)$，$[a_1,+\infty)$，$(a_1,+\infty)$分别为无界开或闭区间。集合$\mathbf{R}$本身也是一个区间，亦即$(-\infty,+\infty)$。

$\mathbf{R}$中所有闭区间的集合用$I(\mathbf{R})$表示，其中每一个区间用大写字母A，B，

$C,\cdots$ 表示，而实数可以看成为点区间，用小写字母 $a,b,c,\cdots$ 表示。

【定义 2.1】 区间数 A 定义为实数 x 的集合，且 $a_1 \leqslant x \leqslant a_2$，即 $x \in [a_1, a_2]$ 或者

$$A = [a_1, a_2] = \{x \mid a_1 \leqslant x \leqslant a_2, x \in \mathbf{R}\} \tag{2.2}$$

当 $a_1 = a_2 = a$ 时，由式(2.2)所给出的区间数 A 就退化成一个实数，记为 $a = [a, a]$。

区间数 A 的度量性质的描述：

(1) 宽度

$$w(A) = w[a_1, a_2] = a_2 - a_1 \tag{2.3}$$

(2) 值

$$|A| = |[a_1, a_2]| = \max(|a_1|, |a_2|) = \begin{cases} |a_1| & |a_1| \geqslant |a_2| \\ |a_2| & |a_1| < |a_2| \end{cases} \tag{2.4}$$

(3) 像

$$A^- = [a_1, a_2]^- = [-a_2, -a_1] \tag{2.5}$$

(4) 倒数

$$A^{-1} = [a_1, a_2]^{-1} = \left[\frac{1}{a_2}, \frac{1}{a_1}\right], 0 \notin [a_1, a_2] \tag{2.6}$$

经常用到的区间数 A 之间的关系有包含、等价。下面分别给出它们的定义：

① 区间数包含关系：假定存在两个区间数 $A = [a_1, a_2]$ 与 $B = [b_1, b_2]$，若有 $b_1 < a_1 < a_2 < b_2$，则称这两个区间数的关系为包含关系，记为

$$A \subset B, \text{即}[a_1, a_2] \subset [b_1, b_2] \text{或} B \supset A, \text{即}[b_1, b_2] \supset [a_1, a_2] \tag{2.7}$$

式(2.7)表明，$[a_1, a_2]$ 区间是 $[b_1, b_2]$ 区间中的一个子集，我们称这种包含为区间嵌套性质。

② 区间数等价关系：在两个区间 $A = [a_1, a_2]$ 与 $B = [b_1, b_2]$ 中，若 $a_1 = b_1, a_2 = b_2$，则称 A 与 B 等价，记为 $A = B$。

如果 A 是一个实数，也就是点 $a = [a, a]$，那么它的像 $a^- = [-a, -a] = -a$，倒数为 $a^{-1} = [\frac{1}{a}, -\frac{1}{a}] = \frac{1}{a}$。由此可得，实数运算是区间运算的特例，而区间运算则是实数运算的推广。因而，区间运算比实数运算要复杂得多。

③ 区间数的运算：区间 $A = [a_1, a_2], B = [b_1, b_2]$ 的算术运算及其规则如下。

(a) 区间数的加法

$$A + B = [a_1, a_2] + [b_1, b_2] = [a_1 + b_1, a_2 + b_2] \tag{2.8}$$

(b) 区间数的减法

$$A - B = [a_1, a_2] - [b_1, b_2] = [a_1 - b_2, a_2 - b_1] \tag{2.9}$$

(c) 区间数的乘法

$$A \cdot B = A \times B = [a_1, a_2] \times [b_1, b_2] = [\min(a_1b_1, a_1b_2, a_2b_1, a_2b_2), \max(a_1b_1, a_1b_2, a_2b_1, a_2b_2)] \tag{2.10}$$

(d) 区间数的除法

$$A \div B = A/B = [a_1, a_2] \div [b_1, b_2] = [a_1, a_2] \cdot \left[\frac{1}{b_2}, \frac{1}{b_1}\right] \tag{2.11}$$

实际上,减法运算 $A - B$ 可以看成是 A 与 B 的像的加法,而除法运算 $A \div B$ 可以看成是 A 与 B 的倒数的乘法。

值得注意的是,$A + A^- = [a_1, a_2] + [-a_2, -a_1] = [-(a_2 - a_1), (a_2 - a_1)] \neq 0$,而 $0A + A^-$,且 $A + A^-$ 的左右两个端点关于 0 是对称的。

$A \cdot A^{-1} = [a_1, a_2] \cdot \left[\frac{1}{a_2}, \frac{1}{a_1}\right] \neq 1$,而 $1 \in A \cdot A^{-1}$。

【例 2.3】　给定两个区间数 $A = [1,3]$,$B = [4,6]$,则

$$A + B = [1,3] + [4,6] = [1+4, 3+6] = [5,9]$$

$$A - B = [1,3] - [4,6] = [1-6, 3-4] = [-5, -1]$$

$$A \cdot B = [1,3] \cdot [4,6] = [\min(1\cdot 4, 1\cdot 6, 3\cdot 4, 3\cdot 6), \max(1\cdot 4, 1\cdot 6, 3\cdot 4, 3\cdot 6)] = [4,18]$$

$$\frac{A}{B} = \frac{[1,3]}{[4,6]} = [1,3] \cdot \left[\frac{1}{4}, \frac{1}{6}\right] = \left[\frac{1}{6}, \frac{3}{4}\right]$$

上述结果显示,如果不确定数 x 与 y 分别位于区间 A 与 B 中,那么 $x \pm y$,xy,$\frac{x}{y}$ 仍然是一个不确定的数,而且它们分别在 $A \pm B$,AB 与 $\frac{A}{B}$ 之中。

区间数运算律

(1) 交换律:$A + B = B + A$,$AB = BA$。

(2) 结合律:$(A + B) + C = A + (B + C)$,$(AB)C = A(BC)$。

点区间 $0 = [0,0]$ 与 $1 = [1,1]$ 对区间加法及乘法而言,是两个唯一的中性元素,具有

$$A = A + 0 = 0 + A, A = A \cdot 1 = 1 \cdot A$$

如果 $a = [a, a]$，那么有下面的运算律。

(3) 分配律：$a(B + C) = aB + aC$。

若用区间数 A 代替实数 a，则上述分配律不是总成立。

模糊数

【定义 2.2】 模糊数

设 U 表示论域，令 $\{A_1, A_2, \cdots, A_n\}$ 表示 U 的子集。u 表示一个对应到 $[0,1]$ 间的实数函数，即 $u: U \to [0,1]$。假定论域 U 上存在一个陈述 X，其相对于子集的隶属函数以 $\{\mu_1(X), \mu_2(X), \cdots, \mu_n(X)\}$ 表示，则在离散情况(Discrete Case)下，陈述 X 的模糊数可表示成

$$\mu_U(X) = \frac{\mu_1(X)}{A_1} + \frac{\mu_2(X)}{A_2} + \cdots + \frac{\mu_n(X)}{A_n} \tag{2.12}$$

其中，+ 表示衔接符号，并不表示加法含义，$\frac{u_i(X)}{A_i}$ 表示陈述 X 隶属于子集 A_i 的程度。当 U 为连续时，陈述 X 的模糊数可表示成

$$\mu(X) = \int_{x \in X} \frac{u_i(X)}{A_i}. \tag{2.13}$$

【例 2.4】 中学某个学生一天运动时间的模糊数表示：

假定 X 表示哈尔滨市某高级中学学生一天运动几个小时，以模糊数表示为 $\mu_\Omega(X)$，论域 Ω 可视为整数论域，即是运动小时数。设 $\Omega = \{0,1,2,3,4,5\}$，X 表示一天运动时间模糊数的隶属函数为 $\{\mu_0(X) = 0.25, \mu_1(X) = 0.4, \mu_2(X) = 0.2, \mu_3(X) = 0.1, \mu_4(X) = 0.05, \mu_5(X) = 0\}$，则 X 的模糊数可表示为

$$\mu_\Omega(X) = \frac{0.25}{0} + \frac{0.4}{1} + \frac{0.2}{2} + \frac{0.1}{3} + \frac{0.05}{4} + \frac{0}{5}$$

【例 2.5】 哈尔滨市居民对松花江“工业园区排污段”水域恶臭感觉的模糊数表示：

假定 X 表示哈尔滨市沿岸居民对松花江“工业园区排污段”水域恶臭的感觉，并用模糊数表示为 $\mu_U(X)$。假定论域 $U = \{1 = 很严重, 2 = 严重, 3 = 普通, 4 = 轻度, 5 = 无影响\}$。若 X 的感觉隶属函数为 $\{\mu_1(X) = 0.25, \mu_2(X) = 0.6, \mu_3(X) = 0.1, \mu_4(X) = 0.05, \mu_5(X) = 0\}$，模糊数亦可以表示为

$$\mu_U(X) = \frac{0.25}{1} + \frac{0.6}{2} + \frac{0.1}{3} + \frac{0.05}{4} + \frac{0}{5}$$

【例 2.6】　投资回报率的模糊数表示：

假定 X 公司投资报酬率以模糊数表示为 $\mu_U(X)$，论域 U 可视为实数论域，即是投资报酬率。设 $U=\{2\%,5\%,10\%,20\%\}$，且令 X 公司投资报酬率模糊数的隶属函数为 $\{\mu_2(X)=0.2,\mu_5(X)=0.4,\mu_{10}(X)=0.3,\mu_{20}(X)=0.1\}$，则 X 公司投资报酬率的模糊数可表示为

$$\mu_U(X)=\frac{0.2}{2\%}+\frac{0.4}{5\%}+\frac{0.3}{10\%}+\frac{0.1}{20\%}$$

模糊逻辑与语言变量

模糊理论是以模糊逻辑为基础，它把传统数学的二元逻辑（又称二值逻辑）做了延伸及推广，而不再是只有对错或是非二分法。对于元素与集合的关系，古典集合论中元素是否属于集合 A，必须十分明确不容模糊，即 $x\in A$ 或 $x\notin A$ 二者必居其一，且只能居其一。这种逻辑正是所谓的二元逻辑。

然而，人类思维源于对自然现象和社会现象的主观意识及影响，其知识语言也会因本身的主观意识、时间、环境以及分析事情的角度不同而具有模糊性。对与错之间还有不完全对、一点对或不完全错等，是非之间还有有些是，有些非等地带，正所谓的灰色地带与模糊观念。这样就出现多值逻辑情况。多值逻辑虽然突破了二元逻辑的限制，承认真假值具有中间过渡性，但它是通过穷举且界限十分明显地来表示这种中间过渡性。例如，“这里是城市的中心”，“那里是城乡结合部”等等，而这个“中心”与“城乡结合部”都是界线不明显的。因此，传统的二元逻辑及多值逻辑就显得无能为力。而采用模糊逻辑则很容易解决这类问题。

设 A 表示论域 U 上的模糊子集，“a 属于 A”就是一个模糊逻辑。这个命题的值可定义为 A 的隶属函数在 a 上的值。于是，模糊命题“a 属于 A”的值不再是非 0 即 1，而是闭区间 $[0,1]$ 上的任何一个值。该值也不像无限值逻辑那样表示模糊命题“a 属于 A”是真还是假，而是表示模糊命题“a 属于 A”是真的程度。比如，哈尔滨属于大城市的隶属度为 0.83，则模糊命题“哈尔滨是大城市”真的程度是 0.83。

语言变量是扎德提出的一个基本概念。语言变量是一个变量，其取值不是具体的数值，而是自然语言或人工语言的字词与（或）语句的变量。语言变量特别适用于描述那些难以精确量化的复杂现象。例如，影像质量就是一个语言变量，此语言变量的取值为：非常高、很高、高、中、可以、差、很差、非常差等等。比如，年龄被表示为语言变量，则其词集 $T(\chi)$ 是语言值的集合，可以是

$$T(\text{年龄}) = \text{年轻} + \text{年老} + \text{很年轻} + \text{不年轻} + \text{很老} + \text{很年轻很年轻} + \text{相当年轻} + \text{有点年轻} + \cdots$$

其中 T(年龄) 中的每个词都是论域，比如 $U = [1,100]$ 的一个模糊子集的标志。

伴随于语言变量有两个规则：(a) 语法规则，其定义 $T(\chi)$ 中的合式语句；(b) 语义规则，以此可以确定 $T(\chi)$ 中词的含义。若 X 是 $T(\chi)$ 中的词，则其含义(指称意义下) 是 U 的子集。$T(\chi)$ 中的基本词是以基本模糊子集为含义的词(此含义必需事先定义的词)，并作为计算 $T(\chi)$ 中非基本词含义的基础。比如，在式(2.16) 中，基本词是年轻及年老，其含义可以由其一致性函数 $\mu_{\text{年轻}}$ 及 $\mu_{\text{年老}}$ 分别定义。对于非基本词的含义——或者等价地，一致性函数——可以通过使用语义规则计算而得。比如，使用后面式(2.20) 可得

$$\mu_{\text{很年轻}} = (\mu_{\text{年轻}})^2 \tag{2.14}$$

$$\mu_{\text{有点年轻}} = (\mu_{\text{年轻}})^{1/2} \tag{2.15}$$

$$\mu_{\text{不很年轻}} = (1 - \mu_{\text{年轻}})^2 \tag{2.16}$$

图 2.12 中给出这些词的一致性函数的图示。

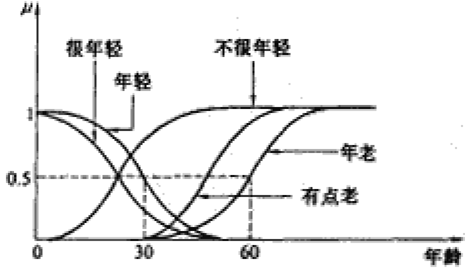


图2.12 年轻、年老、很年轻、有点老、不很年轻的一致性函数(未按比例)

【定义 2.3】 离散模糊数的反模糊化值

设 x 表示离散模糊数，语言变量 $\{L_i \mid i = 1,\cdots,k\}$ 为论域 U 中有序的数列，$\mu_{L_i}(x) = m_i$ 为模糊数 x 相对于语言变量 L_i 的隶属度，$\sum_{i=1}^{k}\mu_{L_i}(x) = 1$，则称 $x_f = \sum_{i=1}^{k} m_i L_i$ 为模糊数 x 的反模糊化值。

【定义 2.4】 连续型模糊数的反模糊化值

设 c 表示连续型模糊数，其隶属函数

$$\mu_U(c)=\begin{cases} f(c) & c\in[a,b] \\ 0 & c\notin[a,b] \end{cases}$$

则 $c_f=\dfrac{\int_a^b cf(c)\mathrm{d}c}{\int_a^b f(c)\mathrm{d}c}$ 为模糊数 c 量化后的值。

在计算模糊数据样本其相对于语言变量的隶属函数时，经常利用三角形隶属函数来进行计算工作，其定义及计算方式如下。

【定义 2.5】　三角形隶属函数的计算

设 U 表示论域，$\{X_i \mid i=1,\cdots,n\}$ 为一组模糊样本。给定 U 的一个有序分割集合，$\{P_j \mid j=1,\cdots,r\}$，且其相对于语言变量为 $\{L_j \mid j=1,\cdots,r\}$。设 m_j 为其分割集合 P_j 的中间值，若 X_i 介于 m_j 与 m_{j+1} 之间，则其属于语言变量 L_j 的隶属度为 $\dfrac{m_{j+1}-X_i}{m_{j+1}-m_j}$，属于语言变量 L_{j+1} 的隶属度为 $\dfrac{X_i-m_{j+1}}{m_{j+1}-m_j}$。

【例 2.7】　计算股市成交股数对应量的语言变量的隶属度。

设 $\{X_i\}=\{16,34,58,70,88\}$（单位：千万）表示上海股市中五只股票在某个月的成交股数，若选择一次序分割集合 $U=\{[0,20],[20,40],[40,60],[60,80],[80,100]\}$，其相对应的语言变量为：微量 $=L_1\propto[0,20)$，小量 $=L_2\propto[20,40)$，普通 $=L_3\propto[40,60)$，大量 $=L_4\propto[60,80)$，巨量 $=L_5\propto[80,100)$，其中"$\propto$"表示"相对于"。再取各分割集合的中间值为 $\{m_1=10,m_2=30,m_3=50,m_4=70,m_5=90\}$，其相对应的语言变量为 $\{L_1,L_2,L_3,L_4,L_5\}$。因为 $X_1=16$ 介于 10 与 30 之间，因此，可计算出 X_1 相对于 L_1 及 L_2 的隶属度如下

$$\frac{30-16}{30-10}=0.7\in L_1,\quad \frac{16-10}{30-10}=0.3\in L_2$$

同理，我们可以得到 $\{X_i\}$ 中每一个元素相对于语言变量的隶属度，如表 2.2 所示。

表 2.2　$\{X_i\}$ 相对于语言变量 $\{L_j\}$ 的隶属度

	L_1	L_2	L_3	L_4	L_5
$X_1=16$	0.7	0.3	0	0	0
$X_2=34$	0	0.8	0.2	0	0
$X_3=58$	0	0	0.6	0.4	0
$X_4=70$	0	0	0	1	0
$X_5=88$	0	0	0	0.1	0.9

[评注]

1. 模糊数据也简称为模糊数，它是模糊集合的元素。

2. 通常，当论域 $U=\mathbf{R}$，把 $\mathbf{R}$ 上的一个模糊集 A 称为模糊数，如果满足：

(1) 对于任意的 $\lambda(0,1)$，A_λ 是 $\mathbf{R}$ 上的有限闭区间，可以表示为

$$A_\lambda=[\underline{a}(\lambda),\overline{a}(\lambda)]$$

(2) 若 A_1 是单点集，即 $A_1=\{a\}$，则称 A 是严格模糊数。

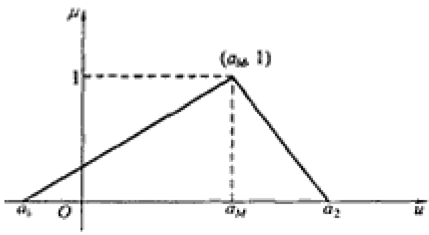
由此可知，一个模糊数的截集族是一个闭区间套。

3. 此外，模糊数还可以被定义为实数域上的、具有凸的分段连续隶属函数的模糊集。最常用的模糊数是三角模糊数与梯形模糊数等，它们都是依据其隶属函数的几何形状命名的。

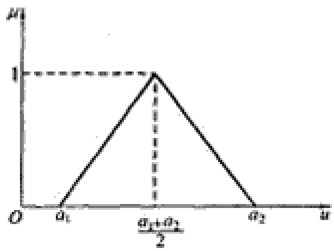
比如，一般的三角模糊数通常可以定义如下。若某一论域 U 上的子集具有

$$\mu(x)=\begin{cases}\dfrac{x-a_1}{a_M-a_1} & a_1\leqslant x\leqslant a_M\\ \dfrac{x-a_2}{a_M-a_2} & a_M<x\leqslant a_2\\ 0 & \text{其他}\end{cases}$$

隶属函数，则称 A 为三角模糊数，其中$[a_1,a_2]$是支撑区间，点$(a_M,1)$是峰值，其图形如图 2.13 所示。对于上式来说，若设 $a_M=(a_1+a_2)/2$，则此时三角模糊数是中心模糊数，用(a_1,a_M,a_2)表示。而设 $a_1=-a,a_2=a,a_M=0$ 时，所得到的三角模糊数关于 μ 轴是对称的，称为对称中心三角模糊数，用$(-a,0,a)$表示。



(a) 三角模糊数



(b) 中心三角模糊数

图 2.13　三角模糊数

总之，各种不同模糊数都可以在模糊集下统一起来。就像整数、自然数、正数等都是数集元素一样。

本节所给出的定义 2.2 是在模糊语言变量下，从模糊数具有三角形隶属函数而加以定义的。本书从后面第 3 章开始研究的主要内容就是这类模糊数。

2.3　模糊关系

在现实世界中，事物是普遍联系的，由于共性而聚集，又因个性不同而相互区别。在普通集合论中的“关系”可以明确地用“有关”或“无关”作出肯定的回答或者否定回答，因此，这样的“关系”抽象地刻画出事物之间的“精确性”联系。然而，“模糊关系”则从更深刻的意义上表现出事物间更广泛的联系。在某种意义上讲，模糊关系的抽象形式更接近于人的思维。

模糊关系是从普通集合中的关系扩展而来，两者之间的实质区别在于用来描述各元素间关系的关系子集的特征函数和隶属度。

普通关系

设 U，V 表示论域，并规定 U 到 V 的一个关系，记为 $U \xrightarrow{R} V$，任意的 $u \in U$ 与 $v \in V$ 之间有这种关系，记为 uRv，或者没有这种关系记为 $u\bar{R}v$，两者必居其一，而且只有这两种状态。

由定义知，一个 U 到 V 的关系，是笛卡儿乘积 $U \times V = \{(u,v) \mid u \in U, v \in V\}$ 的一个子集，亦即 $R \subseteq U \times V$，或者 $R \in P(U \times V)$，对于任意(u,v)用$(u,v) \in R$ 代表 uRv，而$(u,v) \notin R$ 代表 $u\bar{R}v$。

关系 R 的特征函数为

$$X_R(u,v) = R(u,v) = \begin{cases} 1 & uRv \\ 0 & u\bar{R}v \end{cases}$$

模糊关系

在论域 U 与 V 的元素之间，除了有上述那种明确的“有关”或“无关”关系外，还存在着各种程度的模糊关系，比如“相似程度”等关系，显然，这类关系就是笛卡儿乘积 $U \times V$ 的一个模糊子集。

【定义 2.6】　模糊关系

一个 U 到 V 的二元模糊关系 R，是以 $U \times V$ 为论域的一个子集，亦即 $R \in \Phi(U \times V)$。序偶(u,v)的隶属度 $\mu_R(u,v)$ 刻画 u 与 v 之间具有关于 R 的相关程度。

若从映射观点看

$$U \xrightarrow{R} V \begin{cases} \mu_R U \times V \to [0,1] \\ (u,v) \mapsto \mu_R(u,v) \in [0,1] \end{cases} \quad u \in U, v \in V \tag{2.17}$$

在有限论域中，模糊关系通常用矩阵来表示。一般地，对于有限论域 $U = \{u_1, u_2, \cdots, u_n\}$ 到 $V = \{v_1, v_2, \cdots, v_n\}$ 的模糊关系 R 可用 $n \times m$ 矩阵表示为

$$R = (r_{ij})_{n \times m} \tag{2.18}$$

其中，$r_{ij} = \mu_R(u_i, v_j) \in [0,1]$，$i = 1,2,\cdots,n$，$j = 1,2,\cdots,m$，把 R 称为模糊(关系)矩阵。

【例 2.8】 实数域R到R模糊关系"x 远大于 y"：对于任意的 $(x,y) \in \mathbf{R} \times \mathbf{R}$，有

$$\mu_{远大于}(x,y) = \begin{cases} 0 & x \leqslant y \\ \left[1 + \dfrac{100}{(x-y)}\right]^{-1} & x > y \end{cases}$$

经计算得，$\mu_{远大于}(1\,000,100) = 0.999\,9$，$\mu_{远大于}(20,10) = 0.5$。

【例 2.9】 设人的身高论域 $U = \{140,150,160,170,180\}$(cm)，体重论域 $V = \{40,50,60,70,80\}$(kg)。

一个 U 到 V 模糊关系定义为：身高与体重的"标准"相配，将此关系表述成表 2.3。

表 2.3 模糊关系 R

$R(u,v)$ V/kg \ U/cm	40	50	60	70	80
140	1	0.8	0.2	0.1	0
150	0.8	1	0.8	0.2	0.1
160	0.2	0.8	1	0.8	0.2
170	0.1	0.2	0.8	1	0.8
180	0	0.1	0.2	0.8	1

显然，可以将此关系用模糊矩阵表示为

$$R = \begin{bmatrix} 1 & 0.8 & 0.2 & 0.1 & 0 \\ 0.8 & 1 & 0.8 & 0.2 & 0.1 \\ 0.2 & 0.8 & 1 & 0.8 & 0.2 \\ 0.1 & 0.2 & 0.8 & 1 & 0.8 \\ 0 & 0.1 & 0.2 & 0.8 & 1 \end{bmatrix}$$

这种身高与体重间的"标准"关系就可以找出来了。比如，$r_{33} = \mu_R(160,60) = 1$，表示 160 cm 的身高其体重为 60 kg 才算是符合标准，而 $r_{43} = \mu_R(170,60) = 0.8$，表示 170 cm 的身高其体重为 60 kg 还略显得不足。

模糊关系的合成

【定义 2.7】　模糊关系的合成

设 $R \in \Phi(U \times V)$，$Q \in \Phi(V \times W)$，把从 U 到 W 的模糊关系称为 R 与 Q 的合成，记为 $R \circ Q$，其隶属函数为

$$\mu_{R \circ Q}(u,w) = \bigvee_{v \in V}(\mu_R(u,v) \wedge \mu_Q(v,w)) \tag{2.19}$$

特别地，当 $R \in \Phi(U \times U)$，则有模糊关系的幂

$$R^2 = R \circ R, \cdots, R^n = R^{n-1} \circ R \tag{2.20}$$

当论域是有限域时，模糊关系的合成可利用相应的模糊矩阵合成表示。

【定义 2.8】　模糊矩阵的合成

设 $\boldsymbol{R} \in \boldsymbol{M}_{n \times m}$，$\boldsymbol{Q} \in \boldsymbol{M}_{m \times l}$，$\boldsymbol{R} = (r_{ij})_{n \times m}$，$\boldsymbol{Q} = (q_{ij})_{m \times l}$，把从 $\boldsymbol{M}_{n \times m}$ 到 $\boldsymbol{M}_{m \times l}$ 的模糊关系

$$\boldsymbol{S} = \boldsymbol{R} \circ \boldsymbol{Q} \in \boldsymbol{M}_{n \times l} \tag{2.21}$$

称为 $\boldsymbol{R}$ 与 $\boldsymbol{Q}$ 的合成，记为 $\boldsymbol{R} \circ \boldsymbol{Q}$，其中 $\boldsymbol{S} = (S_{ik})_{n \times l}$，$S_{ik} = \bigvee_{j=1}^{m}(r_{ij} \wedge q_{jk})$。

特别地，把 $\boldsymbol{R}$ 与 $\boldsymbol{R}$ 自身的合成矩阵称为模糊矩阵的幂，记为

$$\boldsymbol{R}^2 = \boldsymbol{R} \circ \boldsymbol{R}, \cdots, \boldsymbol{R}^n = \boldsymbol{R}^{n-1} \circ \boldsymbol{R} \tag{2.22}$$

【例 2.10】　设某个家庭的子女与父母的长相相似关系为模糊关系 $\boldsymbol{R}$，其具体表述如下表

R	父	母
子	0.7	0.2
女	0.2	0.6

写成模糊矩阵形式为

$$\boldsymbol{R} = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.6 \end{bmatrix}$$

而父亲与其父′和母′以及母亲与其父″和母″的长相相似关系也是模糊关系，即

R	父′	母′	父″	母″
子	0.7	0.2	0.0	0.0
女	0.0	0.0	0.4	0.7

写成矩阵

$$S = \begin{bmatrix} 0.7 & 0.2 & 0 & 0 \\ 0 & 0 & 0.4 & 0.7 \end{bmatrix}$$

而子女与爷爷奶奶的长相相似关系就是模糊关系 R 与 S 合成，即

$$R \circ S = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.6 \end{bmatrix} \circ \begin{bmatrix} 0.7 & 0.2 & 0 & 0 \\ 0 & 0 & 0.4 & 0.7 \end{bmatrix} =$$

$$\begin{bmatrix} (0.7 \wedge 0.8) \vee (0.1 \wedge 0) & (0.7 \wedge 0.2) \vee (0.2 \wedge 0) & (0.7 \wedge 0) \vee (0.3 \wedge 0.4) & (0.7 \wedge 0) \vee (0.3 \wedge 0.7) \\ (0.3 \wedge 0.8) \vee (0.6 \wedge 0) & (0.3 \wedge 0.2) \vee (0.6 \wedge 0) & (0.3 \wedge 0) \vee (0.6 \wedge 0.4) & (0.3 \wedge 0) \vee (0.6 \wedge 0.7) \end{bmatrix} =$$

$$\begin{bmatrix} 0.7 & 0.2 & 0.3 & 0.3 \\ 0.3 & 0.2 & 0.4 & 0.6 \end{bmatrix}$$

这个合成关系为

$R \circ S$	父′	母′	父″	母″
子	0.7	0.2	0.3	0.3
女	0.3	0.2	0.4	0.6

由此可见，儿子长相像其父亲的父亲，而女儿的长相倒是像其母亲的母亲。

模糊关系合成的性质有：

(1) 结合律

$$(R \circ Q) \circ S = R \circ (Q \circ S)$$

(2) 对并的分配律

$$(R \cup Q) \circ S = (R \circ S) \cup (Q \circ S)$$

$$S \circ (R \cup Q) = (S \circ R) \cup (S \circ Q)$$

(3) 对交的分配律不成立

$$(R \cap Q) \circ S \subseteq (R \circ S) \cap (Q \circ S)$$

$$S \circ (R \cap Q) \subseteq (S \circ R) \cap (S \circ Q)$$

(4)

$$(R \subseteq Q) \Rightarrow S \circ R \subseteq Q \circ S$$

$$R \circ S \subseteq Q \circ S$$

(5)
$$R^n \subseteq Q^n$$
$$R^T \circ Q^T = Q^T \circ R^T$$
$$(R^n)^T = (R^T)^n$$

2.4　模糊等价关系

类似于普通等价关系，模糊等价关系也要满足三个性质：自反性、对称性以及传递性。

【定义 2.9】　模糊等价关系

设 $R \in \Phi(U,U)$，任意的 $u,v \in U$，若满足：

(1) 自反性

$$\mu_R(u,v) = 1 \tag{2.23}$$

(2) 对称性

$$\mu_R(u,v) = \mu_R(v,u) \tag{2.24}$$

(3) 传递性

$$\mu_R(u,w) \geqslant \mu_R(u,v) \wedge \mu_R(v,w) \tag{2.25}$$

则称 R 是 U 上的一个模糊等价关系。

由传递性定义，可以证明，还可使得这个条件更强一些，即传递模糊关系的充要条件为

$$\mu_R(u,w) > \bigvee_{v \in V} \mu_R(u,v) \wedge \mu_R(v,w) \tag{2.26}$$

此外，由合成定义，该条件可写成

$$R \supseteq R \circ R = R^2 \tag{2.27}$$

【定义 2.10】　模糊等价矩阵

对于有限论域情况来说，U 上的模糊等价关系可以表示为一个 $n \times n$ 模糊矩阵 $\boldsymbol{R} = (r_{ij})_{n \times n}$，并满足：

(1) 自反性

$$r_{ij} = 1（也就是主对角线元素为 1）或 \boldsymbol{R} \supset \boldsymbol{I} \tag{2.28}$$

(2) 对称性

$$r_{ij} = r_{ji}（也就是对称矩阵）或 \boldsymbol{R} = \boldsymbol{R} \tag{2.29}$$

(3) 传递性

$$r_{ij} \geqslant \bigvee_{k=1}^{n} (r_{ik} \wedge r_{kj}) 或 \boldsymbol{R} \supseteq \boldsymbol{R} \circ \boldsymbol{R} \tag{2.30}$$

则称此矩阵为模糊等价矩阵。

如果模糊关系 R 仅仅满足(1) 自反性与(2) 对称性，那么称 R 为模糊相似关系。它的隶属度 $\mu_R(u,v)$ 刻画了 u 与 v 关于 R 的相似程度。

类似地，在有限论域情况下，U 上的模糊相似矩阵表现为一个主对角线元素是 1 的对称模糊矩阵。

传递闭包

在实际应用中，利用模糊等价关系或模糊等价矩阵处理问题时，经常获得的是一个具有自反性和对称性的模糊关系(矩阵)—— 模糊相似关系(矩阵)，而传递性却难以满足。然而，可对其加以改进，通过寻找一个包含 R 的最小传递矩阵(也就是传递闭包) 来解决这个问题。

【定义 2.11】　传递闭包

设 $S, S_t, R \in M_{n\times n}, t \in T$。

(1) 若 $S \supseteq S \circ S$，则称 S 为模糊传递矩阵。

可以证明，若 S_t 是模糊传递矩阵，则 $\bigcap_{t\in T} S_t$ 也是模糊传递矩阵，即存在着最小的模糊传递矩阵。

(2) 包含的最小模糊传递矩阵叫做 R 的传递闭包，记为 $t(R)$。它满足：

传递性

$$t(R) \supseteq t(R) \circ t(R) \tag{2.31}$$

包含性

$$t(R) \supseteq R \tag{2.32}$$

最小性

$$S \supseteq R, S \supseteq S \circ S \Rightarrow S \supseteq t(R) \tag{2.33}$$

现在的问题是如何求出 $t(R)$。下面介绍求 $t(R)$ 的方法：

① 总存在着一个传递闭包，并且

$$t(R) = R \cup R^2 \cup \cdots \cup R^m \cup \cdots = \bigcup_{k=1}^{\infty} R^k \tag{2.34}$$

②
$$t(R) = \bigcup_{k=1}^{n} R^k \tag{2.35}$$

也就是，求传递闭包只须有限次的运算。

③ 若 R 又是 n 阶的模糊相似矩阵，则存在自然数 $k \leqslant n$，使得 $t(R) = R^k$，而且对于 $m > k$ 时，恒有

$$R^m = R^k \tag{2.36}$$

由此可得，$t(R)=R^k$ 是一个模糊等价矩阵。

前面的③为人们提供一种具体求出 $t(R)=R^k$ 的方法——平方法：从模糊相似矩阵出发，逐次往后计算，即

$$R^2, R^4, \cdots, R^{2k}, \cdots$$

当第一次出现 $R^{2k}=R^k \cdot R^k=R^k$ 时，这个 R^k 就是传递闭包 $t(R)$，亦即要寻找的一个包含的最小模糊等价关系。这里至多求到 $R^{2^k} \geqslant R^n$，则 R^{2^k} 必为传递闭包。由于 $2^k \geqslant n, k \geqslant \log_2 n$，所以至多只须计算 $k=[\log_2 n]+1$ 步得到 $t(R)$，其中 $[\log_2 n]$ 表示不超过 $\log_2 n$ 的最大整数。

【例 2.11】　求下面模糊矩阵

$$R=\begin{bmatrix} 1.0 & 0.3 & 0.5 & 0.9 \\ 0.3 & 1.0 & 0.6 & 0.2 \\ 0.5 & 0.6 & 1.0 & 0.4 \\ 0.9 & 0.2 & 0.4 & 1.0 \end{bmatrix}$$

的传递闭包 $t(R)$。

解　显然 R 是一个模糊矩阵。有

$$R^2=\begin{bmatrix} 1.0 & 0.5 & 0.5 & 0.9 \\ 0.5 & 1.0 & 0.6 & 0.4 \\ 0.5 & 0.6 & 1.0 & 0.5 \\ 0.9 & 0.4 & 0.5 & 1.0 \end{bmatrix}$$

由于 $R^2 \not\subset R$，所以无传递性，要继续计算。有

$$R^4=\begin{bmatrix} 1.0 & 0.5 & 0.5 & 0.9 \\ 0.5 & 1.0 & 0.6 & 0.5 \\ 0.5 & 0.6 & 1.0 & 0.5 \\ 0.9 & 0.5 & 0.5 & 1.0 \end{bmatrix}$$

$$R^8=\begin{bmatrix} 1.0 & 0.5 & 0.5 & 0.9 \\ 0.5 & 1.0 & 0.6 & 0.5 \\ 0.5 & 0.6 & 1.0 & 0.5 \\ 0.9 & 0.5 & 0.5 & 1.0 \end{bmatrix}$$

由于 $R^4=R^8$，得到 $t(R)=R^4$。

2.5　语意计量与相似度

人类语言的意旨在于和世界沟通。通过研究语意，我们将更深入了解这个

世界共识。然而，即使人们使用相同的字眼、文法或语法。要想成功地进行交流却仍嫌不够，我们希望对相同的东西亦要有同样程度意指。也就是说，我们必须具有一个有关语意的协议。因此，如何分析人类的意指与推断，如何测量它真正感觉的程度，这些问题显得越来越重要。

我们借助于语意计量方法来帮助我们理解语言，进而检验我们的理论。若能建立一种语言计算模型，则拥有和世界沟通的有利工具。在本节里，我们将介绍结合语意学概念与扩展知识，提出建立自然语言计算的系统方法。在讨论语言处理时，应该切记语言的正面属性。因为正是这些正面属性使得语言成为足以包容各种纷繁多彩的现象。

人们常常经由大脑中的印象去寻找或辨认相似的事物，此种辨识能力对于较简单明白的事物具有一定水平的正确率，但对于较复杂的事物就需借助其他工具来完成。本节目的主要在于介绍相似性(Similarity)的衡量方法及分析上应注意的概念。

在语意认知的过程中，最重要且最常被应用的概念就是择近原则。所谓择近原则就是说，对于两件事物，如果彼此之间的距离很大，表示此二件事物相似的程度很低，可能为不同分类的事物；但是，若彼此的距离很小，则表示此二件事物相似的程度很高，可能视为同分类的事物。故在评价计量两集合间的相似度，将论域模糊化的过程中，可把距离较近的两元素给予较高的隶属度。比较两个模糊集的近似距离，可用绝对值差和比距离、欧几里得距离、汉明距离、闵可夫斯基距离等距离公式来测量。一般采用汉明距离和欧几里得距离比较多。而根据解决问题的需要，我们亦可构造出一些其他距离公式。

在相似性问题的探讨中，语言变量的概念亦十分重要。例如，考虑年纪这个变量，它不但可以有其数值，比如间隔[0,100]中的任意数，也可以取年轻、很年轻、相当年轻、极年轻、不年轻、不很年轻、不很老也不很年轻等做值，其中若取后者做值者，就称为语言变量。在一般社会科学研究中，语言变量通常采用五等级，例如，调查对执政党的满意程度，就可分为十分满意、满意、无所谓满不满意、不满意、非常不满意等五等级；父子的相像程度亦可以分成极相似、相似、普通、不相似、极不相似等五级来测量。

语言中语词的语意关系与其他观念有许多相似特征。代表性(Typicality)的认定则须靠统计的方法以达到较高共识。分类(Category)的最高代表性项目应是具有每一因素的高代表性项目。一般已知的语词相关是基于以下 3 种形式：(1) 以代表性显示语意相关程度；(2) 与相对关系作比较；(3) 语意联想

(Semantic Association)。

(1) 代表性

分类的项目代表性有观念上的差异。例如，鳟鱼是很典型的概念鱼。飞鱼则较不具有代表性。甲鱼、青蛙则更不具有代表性。这些鱼类的代表性程度就需要依靠问卷调查进行(软)统计分析完成。分类内的元素项目均以隶属函数表示，再计算其综合加权评分。一般地，我们把语意代表性定义为

$$T(o)=\frac{\text{项目 } o \text{ 的隶属函数值}}{\text{分类中项目对语意的最大隶属度}} \tag{2.37}$$

(2) 一般相似度(General Similarity)

因为人类具有关系相似性的判定本能，故应用于语意关系的比较由来已久。对于单一方面的属性来说，可用类比进行。但是，对于比较复杂的多方面属性来说，必须考虑以数位隶属度计算。若 o_i, o_j 为两对象，(H_i, M_i)，(H_j, M_j) 为主辞(Head)与其相对修辞语(Linguistic Henge)的特征集合，若 o_i, o_j 为两对象，则其相似度可以计量为

$$S(o_i,o_j)=\frac{|H_i\cap H_j|+|M_i\cup M_j|}{|H_i\cup H_j|+|M_i\cup M_j|} \tag{2.38}$$

表 2.4 是语词代表性的一个计量事例。

表 2.4　语词代表性的计量

电线	相似度	喜悦	相似度	肯定	相似度
电线	1	喜悦	1	肯定	1
电缆	0.9	欢怡	0.95	绝对	0.95
光纤	0.7	愉快	0.9	会	0.8
半导体	0.7	开心	0.9	很可能	0.8
水管	0.5	高兴	0.9	大概	0.7
麻绳	0.45	甜蜜	0.8	可能	0.7
针线	0.45	幸福	0.8	也许	0.7
铁丝网	0.3	快乐	0.7	应该	0.7
吸管	0.3	感谢	0.6	或许	0.7
拉链	0.25	哀伤	0.1	或许	0.35
木板	0.05	愤怒	0	否定	0

(3) 偏相似度(Partial Similarity)

若我们仅考虑某单一方面的语意相似度，则称此为偏相似度或部分相似度。设 E, F 表示此两对象，则偏相似度计算方法为

$$PS(E,F) = \frac{\sqrt{T(E \mid c) \cdot T(F \mid c)}}{\max\{T(E \mid c) \cdot T(F \mid c), 1 - T(E \mid c), 1 - T(F \mid c)\}} \tag{2.39}$$

其中，c 表示特定某一方面。

【例 2.12】 假定鱼的概念代表性程度为：T(鳟鱼 | 形状) = 0.9，T(飞鱼 | 形状) = 0.7，T(海豚 | 形状) = 0.6，T(青蛙 | 形状) = 0.1。由式(2.39)与统计样本，可计算此形状(鱼)偏相似度，其结果如表 2.5 所示。

表 2.5 鱼形状的偏相似度

项目	鳟鱼	飞鱼	海豚	青蛙
鳟鱼	1	0.8	0.7	0.3
飞鱼		1	0.6	0.2
海豚			1	0.1
青蛙				1

这个事例只提出简要概念和计算，实际过程与更详细结果应该与语言学者以及统计学家等合作完成。

在模糊识别过程中，模糊熵是类似于热力学中能量含义的标准单位，但是它却与传统意义上的熵又不一样。在定义模糊熵时，并没有使用概率(Probability)论的观念，而是使用扎德(1975)提出的可能性理论(Possibility Theory)的概念，即基于模糊系统的不确定推理理论。模糊熵表示模糊集合的平均内部信息量，此信息量可作为对模糊集合所描述的对象，进行分类时的判定标准。

【定义 2.12】 模糊熵

设 A 表示论域 U 中的一个模糊集合，其隶属度表征为 $\{\mu_1, \mu_2, \cdots, \mu_k\}$，把 A 的模糊熵定义为

$$H(A) = -\frac{1}{k}\sum_{i=1}^{k}[\mu_i \ln(\mu_i) + (1-\mu_i)\ln(1-\mu_i)] \tag{2.40}$$

【例 2.13】 设 X_t 表示一个时间数列，$t = 1, \cdots, N$，u_{it} 为时间数列 X_t 中的元素 x_t 对聚类中心 $C_i \in C$ 的隶属度(参看第 7 章 7.1 节与第 10 章 10.4 节)。x_t 对聚类中心 C_i 的模糊熵定义为

$$\delta(x_t) = -\left(\frac{1}{k}\right)\sum_{i=1}^{k}[u_{it}\ln(u_{it}) + (1 - u_{it})\ln(1 - u_{it})] \qquad (2.41)$$

由定义 2.12 知道，模糊熵表示模糊集合的平均内部信息量，此信息量可作为对模糊集合所描述的对象进行分类时的判定标准。因此，将一个时间数列化为起因于内部可能信息量的度量并用模糊熵表示，可用于有效地辨认此时间数列模型是否有结构性改变的发生。此外，利用 t 时间的平均累加模糊熵来观察模糊熵的信息变化情况，并以此作为模型转折分类的标准。

我们可以把平均累加模糊熵定义如下

$$MS\delta(x_t) = \frac{1}{t}\sum_{i=1}^{t}\delta(x_i) \qquad (2.42)$$

通过观察平均累加模糊熵的走势，可以判定此时间数列是否有结构性改变。换句话说，若平均累加模糊熵发生显著的聚类连串变化时，即意味着模型发生结构性变化；反之，若模型无结构性的变化，则平均累加模糊熵的走势相当稳定，而不会有显著的群聚变化。

此外，无论是自然科学研究还是社会科学的研究，时间数列聚类中心的认定相当主观与分歧，若事先没有特定的评估测量尺度，必会增加实证分析的困难度。因此，模糊分类通常会设定一个 λ 门限水平(Threshold Level)，$0 < \lambda < 1$，来决定分类的标准。

第3章　模糊数据描述统计量

对社会科学研究而言，有时利用模糊数学方法与模型来表示可能比直接给定单一的特定值来得更贴切些，并可能比较适合于用来评估变量与变量之间的相关性。此外，由于其他因素往往也会影响人们对某一事物的评定，因此使用模糊逻辑时，必须对所谓的"其他因素"加以说明，以便把人们的喜好程度转换成便于计算的隶属函数。

对模糊现象进行适当的集合描述就必须使用新的逻辑假设，即论域上的对象由属于某一集合到不属于某一集合是渐进而非突变的。因此，在模糊数据统计分析与检验上，隶属函数的应用起着相当重要的作用。

统计分析在各个领域中得到了广泛的应用。一些基本的描述统计参数，例如期望值、中位数及众数等，并不因其方法简单而失其重要性。在分析数据时，这些参数能够简单且快速地描述数据的基本结构，其中又以期望值最经常被利用。在知识经济的社会中，多元思维逐渐取代传统二元逻辑的思考与分析方法。过去，使用单一数值的样本来计算期望值的方法，已经渐渐不符合现今复杂多变的人工智能科技时代的需求。尤其是在具有多变性、不确定性以及信息不完整性的金融与经济环境下，过分强调对于数值的运算与数学假设的前提，反而更容易造成与现实环境及条件的背离、甚至是脱节。因而，在进行金融与经济方面问题的研究时，利用软计算方法与模糊数据统计分析将会是一种较为合适的测度方法。

在讨论模糊分析调查之前，我们首先给出模糊随机变量的定义及基本观念。想要利用传统的数学逻辑观念，明确定义模糊随机变量并不容易。此处，仅提供一比较合理与符合经验法则的参考。

一般人在日常生活中常对可能发生的事进行推论。例如，一个决策的过程中，某些重要因素的考虑，都可能是下决策的关键。因此，当我们考虑的因素愈多愈详细，就愈能确信最后所作的决策。模糊数据统计就是考虑人们拥有多重喜好，而以模糊区间与加权模糊区间来进行讨论。但是，因人的认知、喜好程度不同，而具有不定等长的期望区间。

在这一章，我们将给出一些模糊数据统计量的定义，为了本书后面各章节阐述方便起见，经常把各种模糊数据统计量简称为模糊统计量，比如模糊样本

均数、模糊样本众数、还有模糊样本中位数等。

3.1　模糊数据样本均数

【定义3.1】　离散型模糊样本均数

设 U 表示一个论域，令 $L=\{L_1,L_2,\cdots,L_k\}$ 为论域 U 上的 k 个语言变量，$\left\{x_i=\frac{m_{i1}}{L_1}+\frac{m_{i2}}{L_2}+\cdots+\frac{m_{ik}}{L_k}, i=1,2,\cdots,n\right\}$为一组模糊样本（$\sum_{j=1}^{k}m_{ij}=1$），则把模糊样本均数（Fuzzy Sample Mean）定义为

$$F\bar{x}=\frac{\frac{1}{n}\sum_{i=1}^{n}m_{i1}}{L_1}+\frac{\frac{1}{n}\sum_{i=1}^{n}m_{i2}}{L_2}+\cdots+\frac{\frac{1}{n}\sum_{i=1}^{n}m_{ik}}{L_k} \tag{3.1}$$

其中 m_{ij} 为第 i 个样本相对于语言变量 L_j 的隶属度。

【例3.1】　离散型模糊均数应用于商品满意度调查。

一种新上市的商品，商品厂商打算探讨消费者的满意程度，于是在街头邀集5位消费者 A,B,C,D,E 作调查，每位消费者对商品满意度的隶属度如表3.1所示。

表3.1　5位受访者对商品满意度的隶属度选择

满意程度	L_1 很不满意	L_2 不满意	L_3 普通	L_4 满意	L_5 很满意
A	0	0.5	0.5	0	0
B	0	0	0.8	0.2	0
C	0	0.3	0.7	0	0
D	0	0	0	0.9	0.1
E	0	0	0.2	0.8	0

则模糊样本均数为

$$F\bar{x}=\frac{\frac{1}{5}(0+0+0+0+0)}{\text{很不满意}}+\frac{\frac{1}{5}(0.5+0+0.3+0+0)}{\text{不满意}}+\frac{\frac{1}{5}(0.5+0.8+0.7+0+0.2)}{\text{普通}}+\frac{\frac{1}{5}(0+0.2+0+0.9+0.8)}{\text{满意}}+\frac{\frac{1}{5}(0+0+0+0+0.1)}{\text{很满意}}=$$

$$\frac{0}{\text{很不满意}}+\frac{0.16}{\text{不满意}}+\frac{0.44}{\text{普通}}+\frac{0.38}{\text{满意}}+\frac{0.02}{\text{很满意}}$$

此模糊样本均数所代表的意义为：很满意的隶属度为 0.02，满意的隶属度为 0.38，普通的隶属度为 0.44，不满意的隶属度为 0.16，很不满意的隶属度为 0。此模糊均数是一个模糊数，表现出此商品的平均满意度最可能为普通、其次为满意。

【定义 3.2】 等距尺度离散型模糊样本均数

设 U 表示一个论域，令 $\{L_1, L_2, \cdots, L_k\}$ 为论域 U 上的 k 个等距尺度变量，$\{x_1, x_2, \cdots, x_n\}$ 为一组模糊样本，且每个样本 x_i 相对于变量 L_j 的隶属度为 m_{ij}，其中 $\sum_{j=1}^{k} m_{ij} = 1$。令 M_j 为 L_j 的组中点，若 $F\bar{x} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} M_{ij}M_j \in L_j$，则把模糊样本 $\{x_1, x_2, \cdots, x_n\}$ 的模糊样本均数定义为 $F\bar{x} = L_k$。

【例 3.2】 人力雇用数量的模糊样本均数。

哈尔滨某机构于近期在上海市拟定成立新办事处，对于规模的大小欲拟定人力雇用计划，于是召集 A, B, C, D, E 5 位相关主管进行意见调查。5 位专家根据给定的人数选项所作出的选择及个别的隶属度如表 3.2 所示。

表 3.2 5 位主管对应于各选项的隶属度选择

人数	1 ~ 3	4 ~ 6	7 ~ 9	10 ~ 12	13 ~ 15	16 ~ 18	19 ~ 21	22 ~ 24
A	0.4	0.6	0	0	0	0	0	0
B	0	0	0.3	0.7	0	0	0	0
C	0	0	0	0	0.7	0.3	0	0
D	0	0	0	0	0	0	0.2	0.8
E	0	0.4	0.6	0	0	0	0	0

由于这 5 位主管有各自的观念及考虑，造成对雇用人数的差异。若要从此样本中得到雇用人数，又要忠实地反映样本的信息，那么使用模糊样本均数是不错的方法。计算如下：令 M_j 为人数区间的组中点 $\{2,5,8,11,14,17,20,23\}$

$$F\bar{x} = (0.4 \cdot 2 + 0.6 \cdot 5 + 0.3 \cdot 8 + 0.7 \cdot 11 + 0.7 \cdot 14 + 0.3 \cdot 17 + 0.2 \cdot 20 + 0.8 \cdot 23 + 0.4 \cdot 5 + 0.6 \cdot 8) \cdot \frac{1}{5} = \frac{58}{5} = 11.6 \in [10,12]$$

故模糊样本均数 $F\bar{x} = [10,12]$。

由此可得，人力雇用人数的模糊样本均数为 [10,12] 这个区间。也就是说，此机构于近期将成立新办事处的雇用人数，参考 5 位主管的意见之后，可以作出平均应该雇用 10 至 12 人的决策。

【定义 3.3】 连续型模糊样本均数（样本为连续型且均匀分布）

设 U 表示一个论域，令 $L = \{L_1, L_2, \cdots, L_k\}$ 为论域 U 上的 k 个语言变量，

$\{x_i = [a_i, b_i], i = 1, \cdots, n\}$ 为论域 U 里的一组模糊样本，则把模糊样本均数定义为

$$F\bar{x} = \left[\frac{1}{n}\sum_{i=1}^{n} a_i, \frac{1}{n}\sum_{i=1}^{n} b_i\right] \tag{3.2}$$

【例 3.3】　哈尔滨某大学对 2006 年毕业生求职潮中，调查出下列 5 位材料研究所毕业的博士生对薪水期望的一组模糊样本为[2 万元,3 万元],[3 万元,4 万元],[4 万元,6 万元],[5 万元,8 万元],[4 万元,7 万元],则根据定义 3.3,其模糊样本均数

$$F\bar{x} = \left[\frac{2+3+4+5+4}{5}, \frac{3+4+6+8+7}{5}\right] = [3.6, 5.6]\text{ 万元}$$

这个信息能提供给需求博士人才的单位主管参考，以便了解目前材料研究所博士毕业生所希望的薪水。

3.2　模糊数据样本众数

【定义 3.4】　离散型模糊样本众数

设 U 表示一个论域，令 $L = \{L_1, L_2, \cdots, L_k\}$ 为论域 U 上的 k 个语言变量，$\left\{x_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}, i = 1,2,\cdots,n\right\}$为一组模糊样本（$\sum_{j=1}^{k} m_{ij} = 1$）。令 $T_j = \sum_{i=1}^{n} m_{ij}$，则称拥有最大的 T_j 值的 L_j 为模糊样本众数（Fuzzy Sample Mode，简记为 Fmode），即

$$Fmode = \{L_j \mid \text{相对于 } j \text{ 项，使得 } T_j = \max_{j=1,2,\cdots,k} T_j\} \tag{3.3}$$

假若存在两组以上的 L_j 其最大 T_j 值相同，则我们称此组数据具有多个模糊样本众数或是具有多重共识。

【推论 1】　设 U 表示一个论域，令 $L = \{L_1, L_2, \cdots, L_k\}$ 为论域 U 上的 k 个语言变量，$\{S_i \mid i = 1, \cdots, n\}$ 为一组模糊样本，且对每个样本 S_i 相对于语言变量 L_j 给予一标准化的隶属度 m_{ij}（$\sum m_{ij} = 1$）。在显著水平 α 下，令 $I_{ij} = 1$，若 $m_{ij} \geq \alpha$；$I_{ij} = 0$，若 $m_{ij} < \alpha$，并且令 $T_j = \sum_{i=1}^{n} I_{ij}$，则称拥有最大的 T_j 值的 L_j 为在显著水平 α 下模糊样本众数，即 $FM = \{L_j \mid \text{相对于 } j \text{ 项，使得 } T_j = \max_{j=1,2,\cdots,k} T_j\}$。假若存在两组以上的 L_j 其最大 T_j 值相同，则我们称此组数据具有多个模糊样本众数或是具有多重共识。

【例 3.4】　利用离散型模糊众数决定旅游地点。

假定哈尔滨市某个科技公司有 10 个人，计划利用周末双休日到户外郊游，

旅游选择地点有太阳岛、植物园、松峰山与二龙山四处。表 3.3 为两种问卷结果的比较。

表 3.3 旅游选择地点两种问卷结果的比较

模糊众数选择旅游地点					传统众数选择旅游地点				
投票	太阳岛	植物园	松峰山	二龙山	投票	太阳岛	植物园	松峰山	二龙山
1	0.4	0.6	0	0	1		○		
2	0.5	0	0.4	0.1	2	○			
3	0.1	0	0.4	0.5	3				○
4	0.4	0	0.6	0	4			○	
5	0	0.8	0.2	0	5		○		
6	0.4	0	0.6	0	6			○	
7	0	0.6	0.4	0	7		○		
8	0.5	0	0.4	0.1	8	○			
9	0.4	0.6	0	0	9		○		
10	0	0	0.7	0.3	10			○	
总计	2.7	2.6	3.7	1.0	总计	2	4	3	1

比较模糊众数与传统众数所决定的旅游地点可以发现，若用传统计票，统计得票结果为植物园 4 票、松峰山 3 票、太阳岛 2 票、二龙山 1 票，则旅游地点众数应为植物园，因为植物园得 4 票最高。

若由模糊隶属度投票计算，则以松峰山隶属度和为 3.7 最高，即选择模糊众数松峰山为旅游地点。但是，选择植物园的 4 票是否就足以代表这 10 个人的最佳共识呢？严格地说，植物园应该只是在二元逻辑的规则下，利用传统众数所求得的偏共识(Partial Common Agreement)。然而，模糊众数较传统众数更能表现出民意之所在，且能找出一个令大家都可接受并且较不极端的结果。

【例 3.5】 哈尔滨市某年中考题库试题难度预估。

在编制题库试题过程中，邀请有关学科及教育测验专家根据其专业，进行选题并预估其对学生而言的难易隶属度 $P(0 \leqslant P \leqslant 1)$。今选取 12 位专家以 $S_1, S_2, \cdots, S_{12}$ 表示，请他们凭借着专业与经验，预估某一试题难易隶属度，利用 $\{L_1, L_2, L_3, L_4, L_5\}$ 分别表示{极难、难、中等、容易、极容易}。这 12 位专家对该题难易度预测的隶属度，如表 3.4 所示。

表 3.4　试题难易度预测的隶属度

专家 \ 语言变量难度	L_1 = 极难	L_2 = 难	L_3 = 中等	L_4 = 容易	L_5 = 极容易
1	0.5	0.4	0.1		
2			0.4	0.6	
3		0.6	0.4		
4		0.4	0.6		
5	0.4	0.6			
6				0.1	0.9
7			0.4	0.6	
8			0.4	0.6	
9		0.4	0.6		
10			0.1	0.1	0.8
11	0.6	0.4			
12			0.4	0.6	
总计	1.5	2.8	3.4	2.6	1.7

若采用传统 5 等级计量表调查，则发现预测情况，如下表 3.5 所示。

表 3.5　试题难易度的预测

专家 \ 语言变量难度	L_1 = 极难	L_2 = 难	L_3 = 中等	L_4 = 容易	L_5 = 极容易
1	○				
2				○	
3		○			
4			○		
5		○			
6					○
7				○	
8				○	
9			○		
10					○
11	○				
12				○	
总计	2	2	2	4	2

利用传统众数方法得知，对该题难度预测的结果为"容易"的教育测验专家最多(2,7,8,12四票)。但是，在12位专家中，仅采用其中4位专家所预测的结果，似乎不太合理。若我们利用模糊隶属度计算，则该题难度的预测结果，为"中等"隶属度总和3.4远大于"容易"隶属度总和2.6。与之相比，利用模糊样本众数会比利用传统众数要来得客观一些，也更能取得大家对该题难度预测的共识。

【定义3.5】 连续型模糊样本众数(样本为连续型且均匀分布)

设 U 表示一个论域，令 $L = \{L_1, L_2, \cdots, L_k\}$ 为论域 U 上的 k 个语言变量，$\{S_i = [a_i, b_i], i = 1,2,\cdots,n\}$ 为论域 U 里的一组模糊样本。若存在一点 x($x \in U$)被样本所覆盖，其覆盖 x 的所有样本为一群落，并且令最大的群落为 MS。则模糊样本众数为最大的群落的所有模糊样本的交集，记为

$$Fmode = [a, b] = \{\cap [a_i, b_i] \mid [a_i, b_i] \subseteq MS\} \tag{3.4}$$

若 $[a, b]$ 不存在(即 $[a, b]$ 为空集合)，则我们称此数据没有模糊样本众数或是无共识；若存在两组以上覆盖频率相同(即 $[a, b]$ 并不唯一)，则我们称此组数据具有多个模糊样本众数或是具有多种共识。

【例3.6】 天津某职业技工学校高级技术工人的就业薪水的模糊样本众数。

设区间 $\{[a_i, b_i] \mid a_i, b_i \in \mathbf{R}, i = 1,2,\cdots,10\}$ 表示10位高级技术工人可接受的薪水范围，单位为万元。$I_1 = [2.5,3.5]$，$I_2 = [2.5,3]$，$I_3 = [3.5,4]$，$I_4 = [2,3]$，$I_5 = [3,4]$，$I_6 = [5,5.5]$，$I_7 = [4,4.5]$，$I_8 = [3,3.5]$，$I_9 = [5.5,6]$，$I_{10} = [2.3,3]$。

根据传统众数的定义，本例无法找出众数，但是，利用模糊样本众数可找出最大群落为 I_1, I_2, I_4, I_{10} 且 $I_1 \cap I_2 \cap I_4 \cap I_{10} = [2.5,3]$。因此，依照定义3.5，25 000至30 000元应为高级技术工人就业薪水的模糊样本众数。

【例3.7】 等候公交车时间的模糊样本众数。

假定我们随机抽取8人，调查他们对于等候公交车时间长短可接受的程度。以区间 $\{[a_i, b_i] \mid a_i, b_i \in \mathbf{R}, i = 1,2,\cdots,8\}$ 表示这8人可接受的等公交车时间，$I_1 = [0,2]$，$I_2 = [9,11]$，$I_3 = [12,15]$，$I_4 = [6,8]$，$I_5 = [3,5]$，$I_6 = [2,3]$，$I_7 = [5,6]$，$I_8 = [8,9]$。

由数据可看出，这8个人可接受的等公交车时间并无交集。故根据定义3.5，我们宣称，此8人对于可接受的等公交车时间并无共识。

模糊样本中位数和传统中位数相同，不会受到样本极端值影响(比如 X_2)，故为一种具有稳健性(Robustness)的估计量，而有关模糊样本极端值的定义则不在本章内容讨论范围之内。

3.3 模糊数据样本中位数

【定义 3.6】 离散型模糊中位数(Fuzzy Sample Median for Discrete Type)

设 U 表示一个论域,令 $L=\{L_1,L_2,\cdots,L_k\}$ 为论域 U 上的 k 个有序变量,$\left\{x_i=\frac{m_{i1}}{L_1}+\frac{m_{i2}}{L_2}+\cdots+\frac{m_{ik}}{L_k}, i=1,2,\cdots,n\right\}$,$\sum_{j=1}^{k}m_{ij}=1$ 为从论域 U 中抽出的一组模糊样本,x_{if} 为对应于模糊样本 x_i 的反模糊化值。

令 $x_{(i)}$ 为 x_i 排序后而得的有序样本值,把离散型模糊样本中位数定义为

$$Fmedian(X)=\begin{cases} x_{(\frac{n+1}{2})} & 若\ n\ 为奇数 \\ (x_{(\frac{n}{2})}+x_{(\frac{n}{2})+1})/2 & 若\ n\ 为偶数 \end{cases} \tag{3.5}$$

【例 3.8】 离散型模糊样本中位数应用于产品定价调查。

某公司欲拟订一种新上市饮料的价格。假定拟订的价格有 20,25,30,35,40(单位:元)5 种方案。公司想探讨何种价格比较能够被消费者所接受,随机抽取 6 位消费者作市场调查,每位消费者对各个价格的隶属度作出评估,结果如表 3.6 所示。

表 3.6　6 位受访者对应于 5 种价格的隶属度选择

价格	20	25	30	35	40	x_{if}
1	0	0	0.5	0.5	0	32.5
2	1	0	0	0	0	20
3	0	0.6	0.4	0	0	27
4	0	0.7	0.3	0	0	26.5
5	0	0	0.4	0.3	0.3	34.5
6	0	0	0.8	0.2	0	31

由于样本数 $n=6$ 为偶数,且 $x_{(3)f}=27$,$x_{(4)f}=31$,而对应 $x_{(3)f}$,$x_{(4)f}$ 的样本值为

$$x_{(3)}=x_3=\frac{0}{20}+\frac{0.6}{25}+\frac{0.4}{30}+\frac{0}{35}+\frac{0}{40}$$

$$x_{(4)}=x_6=\frac{0}{20}+\frac{0}{25}+\frac{0.8}{30}+\frac{0.2}{35}+\frac{0}{40}$$

因而,模糊样本中位数为

$$Fmedian=\frac{0}{20}+\frac{\frac{0.6+0}{2}}{25}+\frac{\frac{0.4+0.8}{2}}{30}+\frac{\frac{0+0.2}{2}}{35}+\frac{0}{40}=$$

$$\frac{0}{20}+\frac{0.3}{25}+\frac{0.6}{30}+\frac{0.1}{35}+\frac{0}{40}$$

若以传统的问卷调查方式，也就是规定每位受访者只能勾选一种意愿最高的选项，则对于受访者而言，所勾选的选项应为心目中隶属度最高者，而其结果如表3.7所示。

表3.7　5位受访者对于5种价格的隶属度选择

价格 / 顾客	20	25	30	35	40
A		○			
B	○				
C			○		
D		○			
E					○
合计	1	2	1	0	1

从上表3.7的数据中我们可以得到，5位受访者的选择价格依大小排序后依序为20,25,25,30,40，而依传统中位数的取法，取出结果为25元。

但我们可以知道，若以传统问卷的方式进行调查，并无法真正反应受访者完整的想法，因为对商品价格问题而言，传统的中位数可能只能以消费者最愿意接受的价格作为考虑，但实际情况却并非如此！若当该商品对消费者而言，十分具有吸引力，即使价格比当初心目中的理想价格要高，消费者仍会有购买的意愿，这样的情况也普遍存在现实生活中，因此，我们若以模糊样本中位数来思考，更能合理分析这类的问题。

从排序后的隶属度中看出，以模糊样本中位数的取出结果为30元，高于以传统中位数取出的25元，从这样结果中，我们可得出结论：以30元的价格来说，一般民众的接受度应不低；而且单价5元多的利润，应是对厂商比较有利的。

【定义3.7】　连续型模糊样本中位数（Fuzzy Sample Median for Continuous Type）

设 U 表示一个论域，令 $\{x_i=[a_i,b_i],i=1,2,\cdots,n\}$ 为从 U 中抽出的一组模糊区间样本。令 c_i 为 x_i 的中点，l_i 为 x_i 的长度。则模糊样本中位数（Fuzzy Sample Median），它以 $median\{c_i\}$ 为中心，以 l_i 的中位数为直径的区间，也就是

$$Fmedian=(c;r)$$

其中

$$c=median\{c_j\},r=\frac{median\{l_i\}}{2} \tag{3.6}$$

【例 3.9】　连续型模糊样本中位数应用于房屋购买意愿调查。

对于商业用户购房屋者来说，一楼店面的面宽对于其营业业绩有很大的影响。通过对数据显示，哈尔滨市某商业街区一楼路旁的房子面宽范围为自 2 m 至 22 m，即论域为 $U = [2,22]$。若某房屋中介公司想了解此街区商业店面面宽需求情况，决定以开放式问卷的形式来了解此街区商业店面面宽实际需求，若随机选取该商业街区 5 位商家作调查，而每位商家对店面面宽的需求区间如表 3.8 所示。

表 3.8　5 位受访者对店面面宽需求的隶属度选择

商家(i)	a_i	b_i	c_i	l_i
1	2	5	3.5	3
2	6	10	8	4
3	7	12	9.5	5
4	3	4	3.5	1
5	17	20	18.5	3

由定义 3.6 可得，模糊样本中位数为(8;1.5)。由模糊样本中位数可知，此街区商家对店面的面宽需求约略为 6.5 m 至 9.5 m 之间，故房屋中介业者可多留意收集此面宽区间的出租或出售案件，提供顾客充足选择以获取最大利润。

【例 3.10】　投资计划金额的模糊样本中位数。

哈尔滨市某公司于近日拟定进行大型投资计划，于是聘请 5 位投资顾问专家进行投资策略分析，5 位专家分别对于投资金额提出建议如表 3.9 所示(金额单位：千万元)。

表 3.9　5 位专家分别对于投资金额提出建议

	A	B	C	D	E
建议投资金额	1 ~ 2	4 ~ 7	3 ~ 8	5 ~ 9	16 ~ 20
c_i	1.5	5.5	5.5	7	18
l_i	2	3	5	4	4

由于 5 个人所提出的建议金额共识度不高，尤其以 E 所提出的投资建议，远大于其他人所建议的金额，但若是直接以平均金额作为投资决策，无疑地会将整体金额提高许多，但是又不能忽略 E 所提的建议，因此，我们考虑以模糊样本中位数的观点来看待分析此问题。

由定义 3.7 可得，区间为(5.5;2) 千万元，此区间即为所求的模糊样本中位

数。

根据以上定义，我们可以对模糊样本进行排序(Ordering)，过程如下：

对离散样本$\{F_i\}$与$\{F_j\}$我们用反模糊化的值来比较大小；而对连续(区间)样本我们则用区间重心的值来比较大小。对于均匀分布的区间样本而言，区间重心即为区间中心，例如若$F_i=[a_i,b_i]$，$F_j=[a_j,b_j]$，如果$\frac{a_i+b_i}{2}\leqslant\frac{a_j+b_j}{2}$时，我们就说$Rank(F_i)\leqslant Rank(F_j)$，而当$\frac{a_i+b_i}{2}=\frac{a_j+b_j}{2}$时，我们称$Rank(F_i)=Rank(F_j)$，即$F_i$与$F_j$形成相持。

3.4 模糊数据统计量的一些性质

统计参数可用来表达总体的特性，但是某些性质却很难借助于传统的统计参数(例如，均数、中位数、众数等表达出来)来刻画。特别是在社会科学研究上，需要确实地表达人们的看法时，显然，传统的统计参数便更不适用了。故本书考虑人类思维的不确定性，根据模糊理论，提出模糊样本众数的定义与分析方法，希望对于人类思维能更完善地表达思想。针对模糊样本众数，我们推导归纳一些有用的性质如下，并比较传统众数与中位数之间的差异，考察它们异同之处。

若以概率统计观点来进行计量研究分析，在数学模型上虽是以繁化简，但却未考虑人类复杂与主观的思维。基于人类的模糊复杂思维，利用模糊统计量来分析数据，是有其深究与探讨的必要性。众数、均数与中位数是统计学常见的三种集中趋势(Central Tendency)测量，它们求法既简单又能清楚表达数据的特性与所传达的信息。因此，若能加入模糊观念利用模糊样本众数、模糊样本均数与模糊样本中位数来从事社会科学的计量研究分析，应是一种实用且重要的方法。

【性质3.1】 设U表示论域，令$\{x_i=[a_i,b_i],i=1,\cdots,n\}$为论域$U$里的一组模糊样本，$L_i$为对应于每一个连续样本$P_i$的区间直径。对连续型模糊样本中位数而言，若超过$[n/2]$个$L_i$等于0时，则模糊样本中位数即为传统中位数。

证明 根据定义3.7，对于连续样本，以传统中位数的取法，所取出的结果应为所有区间组中点的中位数，即是m_i的中位数；而以连续型模糊样本中位数的方法取之时，是以m_i的中位数为中心，所有区间长度L_i的中位数为直径所形成的区间。但是，若超过$[n/2]$个L_i等于0时，显然，L_i的中位数亦为0，则所求出的连续型模糊样本中位数为单一数值，而此即为传统中位数。

【性质3.2】　对于连续型的样本而言，若其样本区间的中心点来自单峰且对称的分布（比如，正态分布），则当样本数足够大时，其模糊样本中位数的中心点将与其模糊样本众数的中心点重合。

证明　由于样本区间的中心点来自单峰且对称分布，当样本数足够大时，所求的模糊样本中位数的中心点应会发生在分布的中点。同样，由于样本区间的中心点来自单峰且对称分布，所以中心点亦会被最多的样本（区间）覆盖，从而形成模糊样本众数的中心点，因此，模糊样本中位数的中心点将与其模糊样本众数的中心点重合。

【性质3.3】　不论离散型或连续型，模糊样本中位数均为存在且唯一，然而，模糊样本众数则未必。

证明　显然，无论样本的分布为何，无论为连续型或是离散型，模糊样本中位数是存在且唯一的。但是，对模糊样本众数而言，如果样本为连续型且呈均匀分布，而若每一个样本均无相交，则没有模糊样本众数的存在。

【性质3.4】　对于离散型的模糊样本中位数而言，可通过改变显著水平，来控制模糊样本中位数里所包含的样本数。

证明　如在例3.2中，原本具有隶属度的选项共有8项，但若考虑删去较小的隶属度，则可以利用选择显著水平 α 的大小作为筛选的标准。若选取显著水平 $\alpha=0.4$，则考虑的样本信息就剩7个；若选取显著水平 $\alpha=0.5$，则考虑的样本信息仅剩5个。由此可知，可通过选择显著水平 α 的大小，来控制离散型模糊样本中位数里所包含的样本数。然而，太大的 α 值却容易造成离散型模糊中数所包含的样本数过少，而降低其代表性；但是，若 α 值取太小，又会造成离散型模糊中数里所包含的样本，有隶属度过低的问题。因此，α 值的选取，需针对不同的问题而有所不同，可凭过去的经验或是所需的样本数，来反推应选取的 α 值。

【性质3.5】　设 $L=\{L_1,L_2,\cdots,L_k\}$ 为论域 U 上的语言变量。若论域 U 上每个样本所给的最大隶属度 m_{ij} 大于显著水平 α 且都落在 L_j 上，则模糊样本众数即为一般众数。

证明　由定义3.4及推论1可得。

【性质3.6】　设 $L=\{L_1,L_2,\cdots,L_k\}$ 为论域 U 上的语言变量，令 $\left\{x_i=\frac{m_{i1}}{L_1}+\frac{m_{i2}}{L_2}+\cdots+\frac{m_{ik}}{L_k},i=1,2,\cdots,n\right\}$ 为论域 U 上一组样本。若存在一群样本，相对于不同的语言变量 $L=\{L_1,L_2,\cdots,L_k\}$ 有相同且最大的隶属度。则在此种情况下，传统众数无法判定。但是，由于 $Fmode=L_j,\max\sum_{i=1}^{n}I_{ij}=1$，

$j = 1,\cdots,k$,若我们选取适当的显著水平 α,当样本 S_i 相对于语言变量 L_j 隶属度 $m_{ij} \geq \alpha$ 时,则可计算模糊样本众数。

证明　由定义 3.4 及推论 1 可得。

【性质3.7】　设 $L = \{L_1, L_2, \cdots, L_k\}$ 为论域 U 上的语言变量,令 $\{x_i = [a_i, b_i], i = 1,\cdots,n\}$ 为论域 U 里的一组模糊样本,m_{ij} 为样本 x_i 对应语言变量 L_j 的隶属度。假若每个样本 S_i 都有一个 $m_{ij} > 0.5$,则对任何显著水平 α,模糊样本众数即为一般众数。

证明　由定义 3.4 及推论 1 可得。

【性质 3.8】　对于连续型的样本而言,若其样本区间的中心点来自单峰且对称的分布(比如,正态分布),则当样本数足够大时,其模糊样本均数的中心点将与其模糊样本中位数的中心点重合。

证明　由于其样本区间的中心点来自单峰且对称的分布(比如,正态分布),则当样本数足够大时,其模糊样本均数的中心点会是整个分布的中点;同样地,模糊样本中位数的中心也会发生在分布的中点。因此,模糊样本均数的中心点将与其模糊样本中位数的中心点重合。

【性质 3.9】　对于连续型的样本而言,若其样本区间的长度来自单峰且对称的分布(比如,正态分布),则当样本数足够大时,其模糊样本均数的区间长度将与其模糊样本中位数的区间长度相同。

证明　由于其样本区间的长度来自单峰且对称的分布(比如,正态分布),则当样本数足够大时,其模糊样本均数的区间长度会是整个分布的中点;同样地,模糊样本中位数的区间长度也会发生在分布的中点。因此,模糊样本均数的区间长度将与其模糊样本中位数的区间长度相同。

【性质 3.10】　对于连续型的样本而言,若其样本区间的中心点及区间长度皆来自单峰且对称的分布(比如,正态分布),则当样本数足够大时,其模糊样本均数等于模糊样本中位数。

证明　由性质 3.1 及性质 3.2 可得:无论区间中心点或区间长度如何,模糊样本均数与模糊样本中位数皆相同,因而,模糊样本均数等于模糊样本中位数。

【性质 3.11】　对于连续型的样本而言,若其样本区间的中心点来自右(左)峰的分布,其模糊样本均数的中心点将大(小)于其模糊样本中位数的中心点。

证明　若其样本区间的中心点来自右(左)峰的分布,其模糊样本均数的中心点会在整个分布的中点偏右(左);但是,模糊样本中位数的中心点会在整

个分布的中点。因此,模糊样本均数的中心点将大(小)于其模糊样本中位数的中心点。

【性质 3.12】　对于连续型的样本而言,若其样本区间的长度来自右(左)峰的分布,其模糊样本均数的区间长度将大(小)于其模糊样本中位数的区间长度。

证明　若其样本区间的长度来自右(左)峰的分布,其模糊样本均数的区间长度会在整个分布的中点偏右(左);但是,模糊样本中位数的区间长度会在整个分布的中点。因此,模糊样本均数的区间长度将大(小)于其模糊样本中位数的区间长度。

【性质 3.13】　不论离散型或连续型,模糊样本均数均为存在且唯一。模糊样本众数则未必。我们可通过调整显著水平 α,来控制模糊样本众数里所包含的样本数。

证明　显然,无论样本的分布为何,连续型或是离散型,模糊样本均数是存在且唯一的。但对于模糊样本众数而言,如果样本为连续型且呈均匀分布,而若每一个样本均无相交,则没有模糊样本众数的存在。

【性质 3.14】　设 $L=\{L_1,L_2,\cdots,L_k\}$ 为离散论域 U 上的(有序)语言变量,令 $\{S_1,S_2,\cdots,S_m\}$ 为论域 U 上一组样本,m_{ij} 为样本 S_i 相对于语言变量 L_j 的隶属度。假定每个样本 S_i 的隶属度 $m_{ij}=1$,则模糊样本均数即为传统均数。

证明　根据定义 3.1,令 m_{ij} 为标准化后样本 S_i 对应 L_j 的隶属度。当样本 S_i 对语言变量 L_j 的隶属度 $m_{ij}=1$ 时,其余 $m_{ik}, k\neq j$ 必定等于 0。

根据传统均数定义,样本 S_i 即是选择语言变量 L_j。而对离散型模糊样本均数,样本 S_i 同样也是相对于语言变量 L_j,故得证。

【性质 3.15】　对于连续型且呈均匀分布的模糊样本均数来说,如果以样本区间的中心点当做样本的确定值,且中心点来自单峰且对称分布(比如,正态分布),那么当样本数足够大时,模糊样本均数会包含传统均数。

证明　因为样本中心点来自单峰且对称分布,由传统均数的定义,当样本数足够大时,均数会发生在分布的中点。因此,中心点亦会被最多的样本(区间)覆盖,而根据定义 3.3,模糊样本均数为

$$F\bar{x}=\left[\frac{1}{n}\sum_{i=1}^{n}a_i,\frac{1}{n}\sum_{i=1}^{n}b_i\right]$$

会包含分布中点,即传统均数。

【性质 3.16】　令 $L=\{L_1,L_2,\cdots,L_k\}$ 为论域 U 上的语言变量,$\{S_1,S_2,\cdots,S_m\}$ 为论域 U 上一组样本。若存在一群样本,相对于不同的语言变量 $L=\{L_1,$

$L_2, \cdots, L_k\}$ 有些语言变量的隶属度不为 1 时，则在这种情况下传统均数无法判定。但是，由于模糊离散型均数，或模糊连续型均数是可以更有效的表现出来，如果我们选取适当的显著水平 α，当样本 S_i 对应语言变量 L_j 的隶属度 $m_{ij} > \alpha$ 时，那么可计算模糊样本均数。

[评注]

当样本对于两个以上的选项具有相同或不同程度的喜好时，传统二元逻辑并不能将此信息表达出来，因此，无法找到均数。然而，我们可以利用隶属度的概念，则将它表达出喜好的程度，并且可根据定义 3.1 及 3.2，求出模糊样本均数。

【性质 3.17】 设 $L = \{L_1, L_2, \cdots, L_k\}$ 为论域 U 上的语言变量，令 $\{S_1, S_2, \cdots, S_m\}$，$\{t_1, t_2, \cdots, t_n\}$ 为论域 U 上二组离散模糊样本，相对于相同的语言变量 $L = \{L_1, L_2, \cdots, L_k\}$，$m_{ij}$ 为模糊样本 S_i 相对于语言变量 L_j 的隶属度，r_{ij} 为模糊样本 T_i 相对于语言变量 L_j 的隶属度，则混合后的模糊样本均数为二组模糊样本均数的平均，与传统的均数意义有所不同。

证明 由定义 3.1 知，$\{S_1, S_2, \cdots, S_m\}$ 的模糊样本均数为

$$F\bar{s} = \frac{\frac{1}{n}\sum_{i=1}^{n} m_{i1}}{L_1} + \frac{\frac{1}{n}\sum_{i=1}^{n} m_{i2}}{L_2} + \cdots + \frac{\frac{1}{n}\sum_{i=1}^{n} m_{ik}}{L_k}$$

$\{t_1, t_2, \cdots, t_n\}$ 的模糊样本均数为

$$F\bar{t} = \frac{\frac{1}{n}\sum_{i=1}^{n} r_{i1}}{L_1} + \frac{\frac{1}{n}\sum_{i=1}^{n} r_{i2}}{L_2} + \cdots + \frac{\frac{1}{n}\sum_{i=1}^{n} r_{ik}}{L_k}$$

又二组相对于语言变量是独立的，如果我们假定 $W_i = \frac{1}{2}(S_i + T_i)$，则 $\frac{m_{ij} + r_{ij}}{2}$ 为样本 W_{ij} 相对于语言变量 L_j 的隶属度。由定义 3.1 可得

$$F\bar{w} = \frac{\frac{1}{2n}\left(\sum_{i=1}^{n} m_{i1} + \sum_{i=1}^{n} r_{i1}\right)}{L_1} + \frac{\frac{1}{2n}\left(\sum_{i=1}^{n} m_{i2} + \sum_{i=1}^{n} r_{i2}\right)}{L_2} + \cdots + \frac{\frac{1}{n}\sum_{i=1}^{n} m_{ik} + \sum_{i=1}^{n} r_{ik}}{L_k}$$

此与传统的均数 $E(Z) = E(X) + E(Y)$ 意义有所不同。

将模糊数据统计分析运用在调查领域来测算的思想，并不受限于特定理论。当客观性伴随着主观性时，系统就会变得模糊。判断一个系统是否为模糊的；也就是说，判断系统是由一确定值或一模糊区间所构成，确有其必要性。

通常，利用确定数值所作的传统抽样调查分析的缺点包括：人的思考行为本来充满着模糊观念及模糊过程，传统问卷的数字经常被过度解释；为迎合数

字的精确需求，实验数据常常有被过度使用之嫌；为了简化或降低数学模型的复杂性，而把实际状况的相关与动态特性忽略。

以模糊区间估计人的喜好，是描绘人们对字句的思维或感觉的一种方法。因此，凡是牵涉到人的思虑问题，利用模糊描述统计量概念来处理是比较合适的。

第4章　模糊数据问卷调查

抽样调查是依据特定研究目的，从人类科学实验或者现实社会中搜集数据以验证事实的一套合乎科学的研究方法。它探讨人类行为模式与思想理念，提供决策参考或作为学术理论基础。但是，在科学研究中所提假设、定义范围及其发展成可测量的经验指标时，由于问卷设计上遣词用句、语意顺序以及问题形式关系，结果就产生相当差异。

传统问卷设计与分析，经常使用勾选的方式让受访者在数个不同选项中选取唯一答案。问题是受访者的回答不只一个时，单一选答结果，可能造成数据难以准确地反映事实真相。本章内容就是利用模糊逻辑概念来改进传统的抽样问卷，允许受访者在数个不同选项中填写隶属度，从而弥补以往传统问卷调查的不足，挖掘人们认识事物思维中的不确定性和多维性的特性。

4.1　社会思维的分歧性与模糊性

社会科学中的调查研究过程，探讨社会、文化、政治、经济、医药、工程、教育和心理等问题，经常以受访者填答问卷(Questionnaire)，通过面对面、电话访谈、田野观测，经由访员填写纪录问卷，作为重要研究的数据依据。

传统的调查过程中，科学家使用某些固定回答模式，测量喜好感觉或满意度，并以整数所衍生的名义尺度记录此评估计量和行为结果间的统计相关性。然而，社会科学中所涉及的概念或情报常常是模糊的，而非确定的、清楚的。比如，在一个典型的调查中，受访者被问及去年一年中的总收入是多少；又比方说，受访者被问及“理想子女数”，“支持政党程度”以及“老人的年龄标准和身份”等问题时，这些问题并没有清楚的界定，受访者在面对这些问题时可能会产生许多不同的答案。

但是，当从事社会经验性研究时，上述的许多问题，经常以某些固定的认定作为回答问题的依据。例如，老人的年龄和身份认定就是一个相当有趣且值得探索的问题。65岁以上是否就可以算是接受老人年金的年龄，或是更年轻一点，或是更年长一些，或是在60岁到70岁之间。但是，若依据老人身体及家庭收

人状况而定，就是一个相当有趣的政策问题。可是，当调查专家设计问卷时，很容易把这种本质上模糊的问题设计成"你赞不赞成政府开办老年年金保险制度？"(中国台湾地区社会意向调查，1994 年 2 月定期调查执行报告)，或"你是否赞成/不赞成 65 岁以上老人接受老人年金" 或"不知道" 作为可能的答项。因此，当问题本身的经验指标可能是模糊的或不一定的状况时，如果用结构性的问卷设计，并且延用清楚的数学模型作为分析的依据，就可能导致很大误差，影响研究结果。

长期以来，在设计问卷时，调查专家们经常感到的一些棘手问题包括：

(1) 有些字词上的解释对于调查员与受访者而言可能会有不同的含义。比如说，在未来几年内这句话，有些人便会认为是指 3 年内，有些人则认为是指 3 到 7 年间，甚至有人会觉得是 7 年以上。由于每个人都有一套自己的感觉，以至于所得到的答案当然会有所偏差。

(2) 有些问项的相关性较强，受访者所回答可能前后不一致，答案可能会在问卷设计者意料之外。

(3) 问项答案选择的多寡，会影响受访者答题时的思考决策。例如，比较是与否的作答选择和具有 5 项选择的作答方式，则结果可能就有差异！

(4) 问项的相对性，也就是说一个问题可能会受到另一个问题的影响。一个很好的例子是，美国在 1950 年代的恐共时期，当受访者被问到说："是否该允许共产党国家的记者入境美国采访？" 时，有 36% 的受访者答："是。" 但是，如果在这问题之前先问："是否认为苏联应让美国记者入境采访？" 时，则回答之前问题的受访者答"是" 的比例从 36% 剧升至 73%。

问题的症结在于，社会调查的设计问卷过程中充满许多模糊概念、模糊框架与模糊归纳逻辑。因为人类的语言、思维与决策，经常具有模糊及不完整的特质，尤其是人类思维模式，更常常出现偏好不确定情况。如果把这些模糊现象用二分法区分的话，可能会获得偏差的结果。为了克服以上困扰，我们势必要找出一些比较可行的研究方法。而利用模糊数据统计技术就是一个很好的解决工具。

在这一章，我们介绍模糊数据统计调查与分析方法，从社会科学的角度来分析人类思考与行为模式。并将这种模糊数据统计与分析方法，实际应用在社会调查中，用以证明此种模式在测量人类模糊思维的可能性及可行性。模糊逻辑的应用，主张个人喜好程度不需要非常清晰或是有条理。因此，模糊逻辑和布尔模式理念是不同的。例如，布尔模式中，A 和 B 比较只有三种结果：$A > B$，

$A < B, A = B$。但是，人类思维是复杂的，兼具有不明确性的喜好。因此，如何对真实状况作出整体描绘的构想，早已是模糊逻辑学者的重要目标。

换句话说，对人类而言，模糊模型比直接指定单一物体一个值，更适合于评估物体间的相关特性。由于其他特性有助于意见的评定，因此使用模糊逻辑时，必须对其他特性也加以说明，以便把人们的喜好程度转换成易于计算的效用函数(Utility Function)。对那些人在思考认知不易表达完善的问题，借助于隶属函数与模糊数据统计分析似乎可提供更明确的表达。本章建立几个重要的模糊数据测量表定义，诸如单项模糊数据测量表、综合模糊数据测量表等等。最后，给出几个实际调查实例，希望通过模糊数据问卷调查及模糊数据统计分析，对人在思想认知不易表达的问题上能够有更合理的测度与解释。

另外，我们也比较应用模糊数据统计所填答之问卷与应用传统方式所填答之问卷当中的差异，并且借助于模糊数据统计所填答的方式与测量技术，改进现有的调查方式，探知人类内心更真实的感受和变化。

4.2 模糊数据问卷设计与特征获取

人的感觉是模糊的。当有人说，今天天气有点冷时，究竟他对于"冷"的定义为何？多少度的温度范围可以称为"冷"呢？对于这样的问题，每个人的回答皆因其主观性而有不同，即使回答者为同一人，也会因为季节、所处的环境不同而有与之前相异的界定范围。当有人问，你今天的心情好不好时，人的心情是否可以用好(1)或不好(0)来完全分隔清楚呢？有时，我们虽肯定地回答是或不是，但其实绝大多数人的内心是介于中间，两者都不是的状态。人类在作判断时，心情经常摇摆不定，而且要明确决定是或不是，要费相当多的精力，因为在最自然的状态下，答案几乎都是模糊的。

人的逻辑思考与计算是依据软计算而进行的，因为即使在条件或数据不明确下，依然能够凭直觉来推断研讨。例如，有人建议下雨时，车子容易打滑，所以要减低速度，然而，若只是下一点小雨，不减速也无妨，但若下的是倾盆大雨，不减速就不行了。诸如此类的判断，对人们来说是极自然的，但是要让计算机来推断，就非常麻烦了，因为把数据输入计算机时，如果没有严格地定义，去掉所有的模糊，则计算机就不会起作用。因此，要让计算机作业，先决条件是必须勉强地把所有定义都弄清楚。人最拿手的是图形认识或剖析抽象化事物，而计算机在这方面就不灵了。因而，到目前为止，人总是迁就自己去配合计算机，如果计算机不引进模糊概念，则研究方法将永远受到束缚无法解脱。

从讲求严谨、准确原则的观点来看，模糊理论也许会让人觉得是在开科学倒车，也有人批评说，这是一门冒牌的学问。甚至有许多人主张，既然人类所有的情感皆是模糊的，就根本不该拿来当做数量研究对象。事实上，模糊理论并不只是与人类本身有关而已，以往那些严密的技术之所以会陷入僵局，是因为若要为该复杂的系统下严密的定义与叙述，需要耗费大量的时间与精力来使其程序化，且辛苦写出的程序，也常因为太过复杂而无法求出答案。然而，这时候只需要了解其大体上的动作，执行大略性的处理手法，即可处理该系统的问题。

除此之外，面对尚存有许多未知部分的系统时该如何解决呢？如果依照传统理论，将"未知部分"视为毫无关联而不予考虑，或是假设一个适当的值，将其程序化。这样一来，那些程序化的数据，无论怎么解析尝试，还是无法掌握到其真正的形式，甚至得出与本质完全无关的结论。因而，未知的部分还是应该当做未知的部分处理。虽然目前尚未找到妥善的方法来解决，但是，我们已经体会到模糊的必要性了。

刘和宋（Liu W.Y and Song N，2001）发展出一种语意近似（Semantic Proximity）的度量方法，基于语意近似的概念，提出模糊关联度（Fuzzy Association Degree）的计算公式，可以比较两个模糊语意集合间的距离。在其公式的计算和证明中，可知其意义是合理且有效的。文中并以植物学的数据分析为例说明，分析稀有植物在生态学上分类的近似程度。卡尔森和富勒（Carlsson and Fuller，2000a，2000b），蒋、周和王（Chiang，Chow，and Wang，2000），赫雷拉，赫雷拉－维德马和马丁内斯（Herrera，Herrera-Viedma，and Martínez，2000），杜波依斯和普拉迪（Dubois and Prade，1991）对于模糊语意的演算方面进行论述，也值得在实际应用方面加以推广。

在分析数据时，一些基本的统计参数，比如期望值（Expected Value）、中位数（Median）以及众数（Mode）等，因计算简单而且又能清楚快速地描述数据本身的基本结构与特性，因此，在许多学科领域中广泛地被应用。然而，传统统计中的众数与期望值，没有考虑到人类思维的模糊性与主观意识的复杂性，于是经常会造成数据无法准确地反映事实真相，因此，本章研究将利用模糊数据统计中的模糊数据众数（Fuzzy Mode）与模糊数据期望值（Fuzzy Expected Value）作为分析数据的工具，以期对人类思维与主观意识作出更合理且精确的描述。

林原宏、郑舜仁、吴柏林（2003）将模糊数据众数应用在教育与心理测验的运用上，利用模糊数据众数来评选教科书、多元评估计量以及九年一贯课程的小中学各年级数学课程时数的安排，结果亦发现，利用模糊众数来反映人们的

想法应是较好的选择。

在问卷调查中，若想要由喜好程度的信息得知个人的理念，可以分别以传统的调查表得到受访者行为意图的点估计，另由模糊问题方式得到模糊统计估计，以便两者加以比较。例如，探讨的问题为：

"在现代社会中(比如中国台湾地区)你预期男性和女性应受几年教育？

(请给一个确定的年数)"

要求受访者回答一个确定的数字，此为传统抽样调查的问题形式。而调查表中另一问题询问：

"在现代社会中(比如中国台湾地区)你预期男性和女性应受几年教育？

(请给一个范围)"

这样的问题要求受访者给定一个区间来回答模糊形式的问题。

以上所介绍的是获得模糊数据以及传统数据的问卷形式。至于属性数据的搜集，传统的问法是要求受访者回答一个确定的等距尺度，属单一逻辑。例如，调查"哈尔滨市太阳岛风景区的游客满意程度"，传统问卷形式为

	1 = 很不满意	2 = 不满意	3 = 普通	4 = 满意	5 = 很满意
您对太阳岛风景区的满意度					

而将模糊逻辑、语意与模糊统计所设计的问卷，运用可复选的形式符合实际状况，可能得到较精确的结果。其问答方式为：

要求受访者在_______线段上划下感受范围(区间)，比如：不满意可记为 __/////___；还好有时很满意__////___///__。用如此的回答形式来获得模糊感受程度。若对分布或隶属度有所了解的一些人员而言，我们亦可引导他们依据对事件感觉的隶属度大小而进行填表。例如：

等　　级	1	2	3	4	5	满意度
您对太阳岛风景区的满意度				0.2	0.8	很满意
您对太阳岛风景区的满意度	0.5				0.5	有些很不满意有些满意
您对太阳岛风景区的满意度		0.3	0.7			还好有点不满意

通常，若有较多的样本数据居于中间位置，在统计学上我们称为集中趋势(Central Tendency)。一般以众数或平均数来代表集中趋势的数量，统计上常用代表总体的三种平均数：算数平均数，中位数，众数，来比较传统和模糊调查数据度量结果的异同。

在传统的统计分析中，平均数都是用点来表示，所以又称为点估计值。但是，在模糊数据统计中，由于分析的是模糊区间数据，因此，亦以区间形式来表示数据的集中趋势，而其运算理念的构思，仍源于传统统计的相关定义。另一方面，同一个问题对于同一受访者而言，其依传统问卷形式所回答的确定数值答案，必包含于依模糊问卷形式回答的模糊区间中，因此，这两种形式的数据，本身即具有某一程度的相关性。

模糊逻辑考虑了人类思维的复杂性及行为的不确定性，允许模糊性现象的存在，比较符合实际状况，这一特点弥补了布尔逻辑的不足。把模糊逻辑的观念运用于问卷调查分析中，提供一个新的收集及分析数据的理念，和不同于往常的统计结果展示方法。

此外，作问卷调查时，对有多重感受的受访者而言，传统问法的单一逻辑，可能使其无所适从，并且当询问的问题涉及个人隐私时，受访者很可能拒答。因此，例如在下一节的案例研究中，允许人们拥有多重感受的模糊理论观念，不但可减少废卷率，还可帮助我们侦测出那些具有双重性或危险倾向的学生，这是传统的问卷形式所无法达成的目标。

4.3　模糊数据测量表

【定义 4.1】　单项模糊数据测量表

设 m 表示感受程度的等级量；p_{ij} 为个人在测量表中第 i 题感受程度在第 j 等级的隶属度，则把单项模糊数据指数 I_{SF} 定义为

$$I_{SF}=\frac{1}{m}\left(\sum_{j=1}^{m}jp_{ij}+\frac{\sum_{j=1}^{m}\left|j-\sum_{j=1}^{m}jp_{ij}\right|\times p_{ij}}{k}\right) \tag{4.1}$$

其中 k 表示感受程度等级中 p_{ij} 不为零的个数。

式(4.1) 中，分母的第一项即为传统的度量方法，第二项则可视为模糊样本的加权计分，其中加权计分的理由为：若受试者的感受越离散，代表此选项熵(Entropy) 越高，活动力较强，应予计量加分。

下面，我们用几个特殊情况阐明定义 4.1 中对模糊感受的加权情况。这里采用比较常用的 5 等级测量表为例($m=5$)：

(1) 当受试者选择 5 时，表示该受试者对此道题感受很强，其个人测量表感受百分比为(0,0,0,0,1)，故我们将此受试者的感受程度定为清晰集合“5”，$I_{SF}=1$。

(2) 当受试者选择3时，表示该受试者对此道题感受普通，其个人测量表感受百分比为(0,0,1,0,0)，故我们将此受试者的感受程度定为清晰集合"3"，$I_{SF} = 0.6$。

(3) 当受试者的选择为两极化时，例如，个人测量表感受百分比为(0.5,0,0,0,0.5)，表示该受试者的感受程度有时很强烈，有时没有感受，故此受试者的感受程度定为模糊集合"$0.5I_1 + 0.5I_5$"，其感受程度为4(=3+1)，$I_{SF} = 0.8$。

(4) 当受试者感受程度均匀分布时，例如，个人测量表感受百分比为(0.2,0.2,0.2,0.2,0.2)，则受试者模糊集合为$0.2I_1 + 0.2I_2 + 0.2I_3 + 0.2I_4 + 0.2I_5$，其感受程度为3.48(3+0.24)，$I_{SF} = 0.648$。

【定义4.2】 综合模糊数据测量表

设 AF 表示模糊综合乐观测量表的分布。令 n 为模糊问卷题数，m 为感受程度的等级量，p_{ij} 为个人在测量表中第 i 题感受程度在第 j 等级的隶属度，k 为感受程度等级中 p_{ij} 不为零的个数，把综合模糊数据指数 I_{AF} 定义为

$$I_{AF} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{m}\left(\sum_{j=1}^{m} jp_{ij} + \frac{\sum_{j=1}^{m}\left|j - \sum_{j=1}^{m} jp_{ij}\right| \times p_{ij}}{k}\right)\right] \tag{4.2}$$

式(4.2)亦可写成

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m} jp_{ij} + \frac{\sum_{j=1}^{m}\left|j - \sum_{j=1}^{m} jp_{ij}\right| \times p_{ij}}{m \times k}\right) &= \frac{\sum_{j=1}^{m} j(\sum_{i=1}^{n} p_{ij})}{n \times m} + \\ &\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{m}\left|j - \sum_{j=1}^{m} jp_{ij}\right| \times p_{ij}}{n \times m \times k}\right) = \frac{\sum_{j=1}^{m} js_j}{n \times m} + \\ &\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{m}\left|j - \sum_{j=1}^{m} jp_{ij}\right| \times p_{ij}}{n \times m \times k}\right)\end{aligned} \tag{4.3}$$

其中最后一个等号第一项内的 s_j 表示为第 j 级感受程度的总和，其中 $j = 1,\cdots,m$。

传统问卷统计分析方法是把各个具有决定性影响的因素视为同等重要，换句话说，各个因素具有相同的权重。但是，实际上应该赋予各个因素不同的权重，即各个因素间相异的重要性程度将决定论域中影响因素的权重。比如说，消费者在购买家电时，可能依据自己的特殊需要，个人喜好或经济能力，商品的价格，售后服务，品牌形象，商品的附加功能，商品的保质期限，折扣的多寡或附送

赠品等等因素而有不同选择。又比如，候选人的政治背景、形象、参选的政治见解或选民对政党的认同度等等因素，都将会影响选民的乐观行为。由此可知，不同的因素会有相异影响决策的重要程度，而对于各个因素的重视程度又将会随着个人的喜恶与主观意识有所改变。假若忽视因素权重的存在，或者给予各个因素相同的权重，或者依分析者个人的主观意识给予权重值，则此错误的分析过程将误导事实，因此，依照因素的重要性不同而给予适当的权重是重要且必要的。

然而，权重值又会因个人特质的差异性而有所改变，也就是说不同个体其所赋予因素的权重值也会不尽相同。社会调查的问卷设计，往往会针对所想要了解的事件状态与现实情况设定论域面，以探究受访者的内心感受和想法。由于每个人内心感受到的界面不同，对事件评判的视角不见得一致，因此，受访者考虑论域中各个影响因子的重要程度也将会因人而异。

在多元化的社会组织中，日积月累的知识，不断变迁的环境与瞬息万变的信息等等，都会使得人类的思维并非直线单一化发展，导致人类的行为变得复杂化、多元化、具有模糊特性，并潜在存在着不确定性。因此，决定论域中各影响因素的权数比重，即所谓的模糊权重也就显得相当重要。

因素权值的决定

由于同一种事物具有多种属性，并且受多种因素的影响，因此，在评价事物的过程中，必须对多个相关的因素作综合性的考虑，亦即进行全面性评价。若这种评判过程涉及模糊因素，就称为模糊综合评判。许多文献研究指出，感觉取向一直是当代行为研究的一项中心课题。因此，假设心理调查的问卷设计主题是在预测受试者感觉动向，而受试者的感觉决定反映出其内在感受或价值的特定取向，所以受试者对不同事物的看法或价值观的评定，有可能兼备所代表的属性特征，进一步构成受试者的个人感觉特质。

受试者在抉择乐观时，必须对多个感觉取向的因素进行全面性的评价，综合地整理归纳出自己具有哪些属性特征。然而，由于受试者对各种感觉取向因素的重视程度会随着其个人喜恶与主观意识有所改变，因此，在影响受试者感觉行为的过程中，不同的乐观取向因素也会有着不尽相同的影响程度。因此，研究者一般首先考虑个体模糊权重值的设定。

个体模糊数据权重值的设定过程

步骤 1：首先设定问卷内容的论域面为乐观，乐观取向因素 $A = \{A_1,$

$A_2, \cdots, A_l\}$。

步骤2：受访者衡量乐观取向因素的重要性程度，将 $A = \{A_1, A_2, \cdots, A_l\}$ 依重要性予以排序，其偏好效用(Perfernce Utility)序列 $U_1 > U_2 > \cdots > U_l$ 为偏好递增序列。

步骤3：令 $V_1, V_2, \cdots, V_l$ 为对应偏好序列 $U_1 > U_2 > \cdots > U_l$ 的排序量值。假定有 n 个受访者，而第 j 个受访者就 A_i 排序所对应的 V_i，用 v_{ij} 表示，其中 $i = 1, 2, \cdots, l, j = 1, 2, \cdots, n$。

步骤4：令 SW_{ij} 为第 j 个受访者给予 A_i 的个体模糊权数值

$$SW_{ij} = \frac{v_{ij}}{\sum_{i=1}^{l} v_{ij}} \tag{4.4}$$

总体模糊数据权重的设定过程

步骤1：令 R_i 为 n 个受访者对 A_i 排序后所得的总排序值

$$R_i = \sum_{j=1}^{n} v_{ij}, i = 1, 2, \cdots, l \tag{4.5}$$

步骤2：令 TW_i 为 A_i 的总体权数值

$$TW_i = \frac{R_i}{\sum_{i=1}^{l} R_i} \tag{4.6}$$

案例研究

在采用八项乐观取向并以其因素来决定乐观倾向的5位专家心理调查中，鉴于每位受访者将对八项取向的因素有不同的感受程度，故请专家按重要性程度以偏好递增序列 $U_1 > U_2 > U_3 > U_4 > U_5 > U_6 > U_7 > U_8$ 的方式加以排序。

这里采用八等级评分标准法，若乐观取向 A_i 为最重要的决定因素时，则取 $V_1 = 8$；若乐观取向 A_i 为次重要的决定因素时，则取 $V_2 = 7$；若乐观取向 A_i 为次不重要的决定因素时，则取 $V_3 = 2$；若乐观取向 A_i 为最不重要的决定因素时，取 $V_4 = 1$。若第 j 个受访者视乐观取向 A_i 为次重要的决定因素，则以 $v_{ij} = 7$ 表示。所得信息由表4.1给出。

表 4.1　5 位受访者的乐观取向决定因素

样本＼项目	1	2	3	4	5	6	7	8
1	6	3	8	7	4	1	5	2
2	6	4	7	8	3	1	5	2
3	5	1	8	7	4	2	6	3
4	5	4	7	8	2	3	6	1
5	5	2	8	7	4	1	6	3
单项总分	27	14	38	37	17	8	28	11
权重(/180)	0.15	0.08	0.21	0.21	0.09	0.04	0.16	0.06

由表4.1得知，受访者 j 对因素 $A = \{A_1, A_2, \cdots, A_8\}$，依照不尽相同的重要程度，给予排序值 $\{v_{1j}, v_{2j}, \cdots, v_{8j}\}$。以第一位为例

$$SW_{11} = \frac{v_{11}}{\sum_{i=1}^{8} v_{i1}} = \frac{6}{1+2+\cdots+8} = \frac{6}{36} = 0.17, SW_{21} = \frac{3}{36} = 0.08$$

$$SW_{31} = \frac{8}{36} = 0.22, SW_{41} = \frac{7}{36} = 0.19, SW_{51} = \frac{4}{36} = 0.11$$

$$SW_{61} = \frac{1}{36} = 0.03, SW_{71} = \frac{5}{36} = 0.14, SW_{81} = \frac{2}{36} = 0.06$$

因此，第一位受访者认为感觉取向因素的个体模糊权数为

$$\{SW_{11}, SW_{21}, \cdots, SW_{81}\} = \{0.17, 0.08, 0.22, 0.19, 0.11, 0.03, 0.14, 0.06\}$$

又由 5 位专家填答得知，其各自的总排序值 $\{R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8\}$ 为

$$R_1 = \sum_{j=1}^{n} v_{1j} = 27, R_2 = 14, R_3 = 38$$

$$R_4 = 37, R_5 = 17, R_6 = 8, R_7 = 28, R_8 = 11$$

则各自的总体权数值可以表示为

$$\{TW_1, TW_2, TW_3, TW_4, TW_5, TW_6, TW_7, TW_8\} =$$

$$\{\frac{27}{180} = 0.15, 0.08, 0.21, 0.21, 0.09, 0.04, 0.16, 0.06\}$$

值得注意的是，对于尚未或无法表示意向，态度不定的受访者或是遗漏数据。由于同一种事物具有多种属性，且受多种因素的影响，因此，在评价事物的过程中，必须对多个相关的因素作综合性的考虑，即进行全部评价，若这种评判过程涉及模糊因素，就称为模糊综合评判。

模糊数据权重推断

步骤 1:令 L_{ij} 为总体权数值与个体权数值的相乘积,即

$$L_i = TW_i \cdot SW_j = TW \cdot \sum_{i=1}^{l} SW_{ij} \tag{4.7}$$

步骤 2:假定尚未表示意向或态度不定之 m 个受访者中的第 j 个受访者,其 A_i 的模糊权重值为 W_{ij}

$$W_{ij} = \frac{L_{ij}}{\sum_{i=1}^{l} L_{ij}}, j = 1,2,\cdots,m \tag{4.8}$$

最后,我们可得模糊数据综合乐观测量表加权指数如下。

模糊数据综合乐观测量表加权指数

设 AF 表示模糊综合乐观测量表的分布。令 n 为模糊问卷题数,m 为感受程度的等级量,p_{ij} 为个人在乐观测量表中,第 i 题感受程度在第 j 等级的隶属度,k 为感受程度等级中 p_{ij} 不为零的个数,TW_i 为第 i 题目的总体权数值。把整体模糊数据乐观测量表加权指数 IW_{AF} 定义为

$$IW_{AF} = \sum_{i=1}^{n} TW_i \left[\frac{1}{m} \left(\sum_{j=1}^{m} jp_{ij} + \frac{\sum_{j=1}^{m} \left| j - \sum_{j=1}^{m} jp_{ij} \right| \times p_{ij}}{k} \right) \right] \tag{4.9}$$

样本代表性检验及样本反复加权修正

在问卷调查分析中,抽样所得到的样本偏差,会造成总体特征的错误推断,因此,为了解决此问题,我们可以考虑使用卡方拟合度检验来判断样本能否充分反映总体状况,以增加样本统计推断的精确度。

利用抽样调查数据对总体进行推论时,可以从两方面决定是否对抽样所得到的样本有进行加权的必要。第一,如果数据是来自不等概率抽样,或是分层抽样但各层抽取率不相等,或是等概率抽户而每户访问一人,在这些状况下一定要加权,否则推断即有偏差的。第二,如果数据是来自涵盖性不足或是访问失败的结果,则从逻辑上来说,加权处理适用已经得到的数据去替代那些因各式各样的原因未观察到的数据。也就是,假定在一个相同权值的范围内以取得的数据与应该取得但未取得的资料没有差异。

针对涵盖性不足或是访问失败的部分进行追踪观察,运用统计方法检验无差异假设是否成立,再决定如何处理。同时,如果遇到多个变量的情况,认为多

个变量逐一加权反复操作到收敛为止，需要的总体数据较容易取得，成本低、效果却很大，且每一个考虑到的变项代表性都将通过统计 t 检验。

【例 4.1】　假定有 10 位受访者接受民意调查，以候选人的政治表现、候选人所建立的政治形象、发表的政见、所属的党派为投票取向。以此 4 项投票取向的因素来决定心目中最理想的候选人。但是，鉴于每位受访者将对此处 4 项投票取向的因素有不同的感受程度，请受访者按政治表现、政治形象、政见、所属党派的重要性程度以偏好递增序列 $U_1 > U_2 > U_3 > U_4$ 的方式排序。

这里采用四等级评分标准法，若投票取向 A_i 为最重要的决定因素时，取 $V_1 = 4$；若投票取向 A_i 为次重要的决定因素时，取 $V_2 = 3$；若投票取向 A_i 为次不重要的决定因素时，取 $V_3 = 2$；若投票取向 A_i 为最不重要的决定因素时，取 $V_4 = 1$。若第 j 个受访者视投票取向 A_i 为次重要的决定因素，则以 $v_{ij} = 3$ 表示。所得信息由表 4.2 给出。

由表 4.2 知，第一位受访者对投票取向因素 $A = \{A_1, A_2, A_3, A_4\} = \{$政治表现、政治形象、政见、所属党派$\}$，依照不尽相同的重要程度，给予排序值 $\{v_{1,1}, v_{2,1}, v_{3,1}, v_{4,1}\}$ 为 $\{4,2,3,1\}$。

表 4.2　10 位受访者重要的决定因素投票取向

投票取向受访者	政治表现	政治形象	候选政见	所属党派
01	4	2	3	1
02	4	3	1	2
03	3	4	2	1
04	4	3	2	1
05	4	1	2	3
06	4	2	1	3
07	3	2	4	1
08	4	2	3	1
09	3	4	2	1
10	4	3	2	1

$$SW_{1,1} = \frac{v_{1,1}}{\sum_{i=1}^{4} v_{i,1}} = \frac{4}{4+2+3+1} = 0.4$$

$$SW_{2,1} = \frac{2}{10} = 0.2,\ SW_{3,1} = \frac{3}{10} = 0.3,\ SW_{4,1} = \frac{1}{10} = 0.1$$

即第一位受访者认为投票取向因素{政治表现、政治形象、政见、所属的党派}的个体模糊权数$\{SW_{1,1},SW_{2,1},SW_{3,1},SW_{4,1}\}=\{0.4,0.2,0.3,0.1\}$。

从事民意调查时，常常会遇到不愿表示意向或态度不定的受访者，使得预测的工作不易进行，如何处理这些缺失值(Missing Value)是我们致力解决的目标。一些研究指出，意向不明确的受访者将会受到内在与外在环境的影响，故本章研究除了考虑个体模糊权数值外尚考虑总体模糊权数值。视个体模糊权数值与总体模糊权数值的相乘积为模糊权数值。

【例 4.2】 (例 4.1 续)由 10 位受访者填答中得知，{政治表现、政治形象、政见、所属的党派}各自的总排序值$\{R_1,R_2,R_3,R_4\}$为

$$R_1=v_{1,1}+v_{1,2}+v_{1,3}+\cdots+v_{1,10}=37$$

$$R_2=v_{2,1}+v_{2,2}+v_{2,3}+\cdots+v_{2,10}=26$$

$$R_3=v_{3,1}+v_{3,2}+v_{3,3}+\cdots+v_{3,10}=22$$

$$R_4=v_{4,1}+v_{4,2}+v_{4,3}+\cdots+v_{4,10}=15$$

则{政治表现、政治形象、政见、所属党派}各自的总体权数值可以表示为$\{TW_1,TW_2,TW_3,TW_4\}=\{0.37,0.27,0.21,0.15\}$。

$$TW_1=\frac{R_1}{\sum_{i=1}^{4}R_i}=\frac{37}{100}=0.37,TW_2=\frac{27}{100}=0.27$$

$$TW_3=\frac{21}{100}=0.21,TW_4=\frac{15}{100}=0.15$$

假定第一位受访者不愿明确表示支持哪位候选人或其态度不定，此时应该考虑模糊权数值的设定。

$$L_{1,1}=TW_1\cdot SW_{1,1}=0.37\cdot 0.4=0.148,L_{2,1}=0.27\cdot 0.2=0.054$$

$$L_{3,1}=0.21\cdot 0.3=0.063,L_{4,1}=0.15\cdot 0.1=0.015$$

$$W_{1,1}=\frac{L_{1,1}}{\sum_{i=1}^{l}L_{i,1}}=\frac{0.148}{0.148+0.054+0.063+0.015}=0.528\,6$$

$$W_{2,1}=\frac{0.054}{0.28}=0.192\,8,W_{3,1}=\frac{0.063}{0.28}=0.225,W_{4,1}=\frac{0.015}{0.28}=0.053\,6$$

即第一位受访者对投票取向因素的模糊权数值的设定为

$$\{W_{1,1},W_{2,1},W_{3,1},W_{4,1}\}=\{0.53,0.19,0.23,0.05\}$$

如何估计游离票问题

受访者在作最后的决定时，或许会受外界环境、传播媒体、他人舆论评判、

时间变动等等因素的影响，也因为个人内心思维、行事态度的模糊不确定性，都将给予个人的思考逻辑与作风处事很大的弹性空间，进而表现出个人的最终决策的确定度不够，产生暧昧的、不确定的、游离状态。

在选举的调查研究中，受访者不表示其投票意向的情况非常普遍，因此，预测工作的最主要任务之一，即是在决定那些不表达意见的选民可能的投票意向是什么？不表明投票意向的人，我们对他们并非一无所知。很可能除了投票意向之外，我们很清楚他的社会背景，他对候选人、政党、政治议题、或其他政治标的所具有的态度。这些态度中的相当多项目对投票意向的影响很大，这是我们从以往中外研究的结论中得知的。

我们的研究表明，意向模糊的受访者的决策倾向可以使用模糊统计隶属度概念，来过滤人类本质的模糊性，考虑人们将会以最大隶属度为决策的依据。

最大隶属度判定过程

(1) 令 $T_j(i,k)$ 为第 j 个未决定意向受访者就第 k 个候选人给予第 i 个投票取向因素的评比分数。

$$X_{j,k}=\sum_{i=1}^{n}W_{j,i}\cdot T_j(i,k) \tag{4.10}$$

$$X_j=\sum_{k}X_{j,k} \tag{4.11}$$

(2) 令第 j 个受访者就第 k 个候选人给予的支持隶属度为 $M_{j,k}$

$$M_{j,k}=\frac{X_{j,k}}{X_j},\ \sum_{k}M_{j,k}=1 \tag{4.12}$$

即第 j 个受访者的隶属函数可由 k 位候选人属性相对所属的隶属度 $\{M_{j,1},M_{j,2},\cdots,M_{j,k}\}$ 表示如下

$$\mu_{Aj}(O)=M_{j,1}I_{O_1}(O)+M_{j,2}I_{O_2}(O)+\cdots+M_{j,k}I_{O_k}(O) \tag{4.13}$$

(3) 假定第 j 个受访者会就最大隶属度 M_j 决定最终支持的候选者

$$M_j=\max\{M_{j,1},M_{j,2},\cdots,M_{j,k}\} \tag{4.14}$$

(4) 令 $D_{j,k}$ 为第 j 个受访者最终决定支持第 k 位候选者

$$D_{j,k}=1,\text{若 } M_j=M_{j,k};D_{j,k}=0,\text{其他} \tag{4.15}$$

(5) 若第 k 位候选者在尚未明确表示决定意向的受访者中，则所获得的支持票数为 Y_k

$$Y_k=\sum_{j=1}^{n}D_{j,k}X_j \tag{4.16}$$

最大隶属度判定的可靠性分析

对于已经表明投票意向的受访者，应将其计人候选人的得票。为了测量最大隶属度判定与已经表态受访者之间的信心度，尝试对于已表态的受访者作最大隶属度判定的可靠性分析(Reliability Analysis)，作为最大隶属度预测尚未表态的受访者的动态修正依据，以期比较符合尚未表态的受访者的实际行为。

最大隶属度预测意向依据可靠性分析法尝试动态修正过程

(1) 令 O_i 为第 i 位候选人；TO_i 为第 i 位候选人在已经明确表态的受访者中，所获得的总票数；TMO_i 为第 i 位候选人在尚未明确表态的受访中，依据最大隶属度的判定过程后，所获得的总票数：TO 为第 i 位候选人在此研究中所获得票数的总和。

(2) 在已经明确表态的受访者中投票给第 i 位候选人；且依据最大隶属度的判定过程后，也投票给第 i 位候选人，记为 O_{ii}(例如，参看表 4.3)。在已经明确表态的受访者中投票给第 i 位候选人；但是依据最大隶属度的判定过程后，却投票给第 j 位候选人，记为 O_{ij}。

$$TO_i = \sum_{i=1}^{n} O_{1,i} \tag{4.17}$$

$$TMO_i = \sum_{i=1}^{n} O_{i,1} \tag{4.18}$$

$$TO = \sum_{i=1}^{n} TO_i = \sum_{i=1}^{n} TMO_i \tag{4.19}$$

依据可靠性分析法决定动态修正的原则，如表 4.4 所示，将调整第 k 位候选者在尚未明确表示决定意向的受访者中，所获得的支持票数。

表 4.3 已表态的受访者的投票意愿与最大隶属度判定的综合比较

最大隶属度判定 \ 表态的受访者	O_1	O_2	…	O_n	总得票数
O_1	O_{11}	O_{12}	…	O_{1n}	TO_1
O_2	O_{21}	O_{22}	…	O_{2n}	TO_2
⋮	⋮	⋮		⋮	⋮
O_n	O_{n1}	O_{n2}	…	O_{nn}	TO_n
总得票数	TMO_1	TMO_2	…	TMO_n	TO

表 4.4　依据可靠性分析修正最大隶属度判定的原则表

最大隶属度判定 / 表态的受访者	O_1	O_2	…	O_n	总得票数
O_1	$\frac{O_{11}}{TMO_1}$	$\frac{O_{12}}{TMO_2}$	…	$\frac{O_{1n}}{TMO_n}$	$\frac{O_{11}}{TMO_1}$
O_2	$\frac{O_{21}}{TMO_1}$	$\frac{O_{22}}{TMO_2}$	…	$\frac{O_{2n}}{TMO_n}$	$\frac{O_{21}}{TMO_1}$
⋮	⋮	⋮		⋮	⋮
O_n	$\frac{O_{n1}}{TMO_1}$	$\frac{O_{n2}}{TMO_2}$	…	$\frac{O_{nn}}{TMO_n}$	$\frac{O_{n1}}{TMO_1}$
总得票数	TMO_1	TMO_2	…	TMO_n	TO

$$RF_{k,k} = \frac{O_{k,k}}{TMO_k}, \quad k = 1,2,\cdots,n \tag{4.20}$$

$$RF_{k,h} = \frac{O_{k,h}}{TMO_h}, \quad k = 1,2,\cdots,n, k \neq h \tag{4.21}$$

以最大隶属度预测尚未表态的受访者的决定意向后，再通过可靠性分析修正原则表作为意向预测的动态修正依据，那么第 k 位候选者在尚未明确表示决定意向的受访者中，所获得的支持票数为 T_k

$$T_k = Y_k RF_{k,k} + \sum_{h \neq k} Y_h RF_{k,h} \tag{4.22}$$

近年来，广泛运用民意调查的预测，使得整个社会预测趋势、市场营销等活跃起来。尤其是在选举期间，政党、候选人、民调公司、大众媒体、学术界都热衷于选情预测。假若事前的选情预测能够感受到选民的抉择心态，就可以准确地洞悉选举期间的现状，据此可当做竞选宣传和政党政见调整的依据。

4.4　个案研究：选民投票意向与选情预测

在 1900 年初，美国就开始进行调查选民投票意图，由收集得来的样本信息预测选民投票行为，预测选举结果（吉本斯，Gibbons，1992）。近年来，一些民意调查研究者，关注于预测竞选期间选情动向的流行趋势，试将问卷设计为探讨选民们投票理念与投票行为意图，用于预测选民投票意向。民意调查的分析结果所提供的信息，经常成为社会舆论关注的焦点之一，也将作为党派于选举期间的政治运作的考虑依据，也为候选人欲争取更多选票的参考信息。民意调查的预测结果是否能逼近事实的选举结果，成功的预估每一位候选人的得票率且

正确地判断当选者，都成为研究者努力的方向。

美国密西根大学所发展起来的政党认同、候选人评价、政见立场被最广泛用于研究影响投票决定的因素。但是，根据其他的参考文献可得知，影响投票决定的因素众多，诸如选民所属的年龄层、性别、省籍情结、教育程度、宗教信仰、经济能力、居住的行政区、传播媒体效益、竞选宣传战、选举历史、居住地过去的开票纪录、社会经济指标等等。

选情预测的问卷设计

选情预测的问卷设计，应注意问卷陈述问题的形式，分析问卷所依据的统计理论，应用何种统计分析方式，这些对研究结果的解释有很大的影响。即使相同的问卷题目，也会因不同的问卷设计形式，得到截然不同的结果；而且相同的问卷设计形式，又会随着分析问卷的过程，导致差异极大的结论。这里我们大致可以把问卷设计形式与统计分析方式分为三类：(1) 传统统计分析传统问卷；(2) 模糊统计分析传统问卷；(3) 模糊统计分析模糊问卷。

由于多数研究文献都认为候选人与政党因素将会影响选民的投票行为，所以通常的研究问卷设计选民的投票取向因素设定为{政治表现、政治形象、政见、所属党派}，并把此四种因素列为问卷内容的前四题题目，视此四种因素将会影响候选人所获得的支持度。为了了解受访者对投票取向因素的认同度，设计0至10的固定分数区间，其中0表示最不满意，而10表示最满意，5表示普通，请他们依照感受的喜好程度而自由地为数位候选人作全面性的评比。

然而，选民把自己的选票投给自认为理想的未来某候选人的决策过程中，其抉择所依赖的投票取向因素，未必有着同等重要程度的影响力，也就是选民的投票理念可能依个人的主观看法而特别偏重某因素。因此，设计问卷内容的第五题，请受访者将影响他投票决定的投票取向因素{政治表现、政治形象、政见、所属党派}，依个人的感受和主观看法，就影响的重要程度依次排序。

传统选情预测的问卷设计，都是以单一点的确定数值表示，可分为五等级与语意差别尺度两种形式。前者以均等分数差距转换受访者就问项的答案，比如{非常不同意，不同意，无意见，同意，非常同意} = {1,2,3,4,5}，以顺序性的语意措辞尺度回答方式，将会使得受访者容易填写问项的回答项以表现自己的看法。语意差别尺度，是属于连续型的区间尺度，受访者可依据自己切身的感受选择适合表达态度强弱的位置，但是，语意差别尺度法不易令受访者真正了解如何选择连续性区间尺度。

检视传统问卷在设计上的缺失，最近的问卷调查的回答项，不再仅为二元

逻辑的勾选形态，还包括不知道、两者都不是、稍稍认为、不太认为如此、其他形态的勾选栏等等，以期望更接近人类实际上的思考模式，我们不难发现，这种问卷调查已经是模糊理论的产物。在模糊数据问卷的设计上，考虑了人类复杂的思维与行为模式，应用区间值的模糊特性，估计人类对事物的喜恶强弱。可以依据问卷的设计形式分为：要求受访者在连续区间线段内划出自身感受的范围，或是请受访者回答问项时给予一个感受的连续区间值。

不定等长且可复选的模糊期望区间，将会反映出个人对问项的认知程度，因此，连续的模糊区间值具有处理事态为亦真亦伪的能力。然而，一般受访者可能对模糊问卷的设计方式的接受度不高，经常不能够体会在模糊连续区间中，如何划出线段范围或填写感受区间值。

我们保留传统问卷的优点，也就是使用顺序性的语意措辞尺度，令受访者可以容易地选择适合表达切身感受的回答方式，同时考虑到大多数人的心意，仍介于两者都不是的中间状态，大多数人的内心意向会随着时间的不同、环境情状变化等等，进而改变决策。为了能够更接近人类的思考模式，想通过模糊理论的隶属度数学模型，逼近受访者的思考逻辑。

下面给出的实例，把候选人职务表现、政治形象、候选政见、所属政党列为预测 ABC 市长选举的探究因素，应用模糊统计分析传统问卷，认为每一位选民将因本身的独特因素特征，各视影响选情的论域有着不尽相同重要程度，故考虑个体模糊权重、总体模糊权重、模糊权重数的选定，评定 ABC 市三位候选人，比如 Damer，Eden，Ferry 将各具有不同程度的隶属函数，利用隶属度与决定投票意向之间的微妙关系，以最大隶属度预测选民最终决定的投票意向，将对复杂的选举情势作更适合的预测估计。

但是，受访者有时会因为问卷本身的问题内容太复杂，导致对问题的了解有不同程度的困难存在，或认为问题的敏感性太强具有侵犯性，导致受访者答题的防卫心理，或因为传统问卷调查的二元逻辑不符合个人的思想理念与行为模式，造成选定回答项的困扰，或受访者不愿表示内心意见与态度，倾向社会大众赞赏方向作答等等因素，将导致问卷中无意见或拒绝回答者占相当大的比率，且受访者回答的可靠程度，都将会影响问卷调查的预测结果。如何处理无意见或拒绝回答者的问卷，势必成为研究者努力的方向之一，尤其在选情预测研究方面。

【例 4.3】　假定论域 U 为候选人所得到选民的支持度，候选人的支持度是以{政治表现、政治形象、政见、所属党派}四项影响投票取向因素所决定，而所有候选人的属性 $x \in U$，个别的受访者 $A \subseteq U$。我们给予 0 到 1 之间的数值，表

示所有候选人的属性对个别受访者的隶属程度。假定第一位受访者的隶属函数可由候选人属性相对所属的隶属度{0.4,0.5,0.1}表示

$$\mu_A(x) = 0.4I_{\text{Damer}}(x) + 0.5I_{\text{Eden}}(x) + 0.1I_{\text{Ferry}}(x)$$

若问及可能投票给哪位候选人组合的回答时，比如 Damer 或 Eden、Eden 或 Ferry、Ferry 或 Damer 的组合。第一位受访者的隶属函数，可由候选人属性相对所属的隶属度可能情况如下(尚未经过标准化)

$$\mu_A(x) = 0.9I_{\text{Damer}\cup\text{Eden}}(x) + 0.6I_{\text{Eden}\cup\text{Ferry}}(x) + 0.3I_{\text{Ferry}\cup\text{Damer}}(x)$$

本研究实例以电话访问作为数据收集的依据。应用随机选取电话号码作为抽样所得到的样本。令所得的电话访问样本具有科学合理性，再根据所设定抽样样本的特性，其样本年龄层必须达到法定年龄 20 岁以上，拥有公民投票权，且设置户籍于 ABC 市的选民。1998 年 11 月 23 日至 12 月 3 日为此次电话访问进行期间，其成功样本数为 461 份，抽样误差为 4.66%。

因为考虑问卷数据的收集若以电话访问方式取得，将模糊理论套用在设计问卷题目回答项上，也许访员可能无法完整地传达清受访者以模糊区间表现自己的喜好的信息，或者因为受访者本身意会答题方式的接受程度不高。故如何应用传统问卷的设计形式，尝试套用模糊数据统计观念分析问卷，并期不忽略人类内心的模糊性与不确定性，成为本章努力追求的目标。

样本代表性检验及样本反复加权修正

在问卷调查分析中，成功样本的抽样偏差，将会使得抽样样本无法真正地显示总体特征。为了修正抽样误差，我们考虑使用卡方拟合度检验来判断成功样本的代表性，是否成功样本与总体的人口结构{性别的男女比例、年龄层分配比例、户籍所属行政区的分布比例、教育程度比例}互相符合，足以进行推断总体?是否需要考虑对样本修正加权数，使得样本分布结构与总体特征相符合，以增加样本推断的精确度。

问卷分析过程

461 份有效问卷中，已明确地回答第七题问项的最有可能投给哪位候选人者有 400 份，占总有效问卷的 86.77%；而回答不知道或尚未决定或甚至拒绝回答者有 61 份，占总有效问卷的 13.23%。考虑样本代表性，经过反复加权法修正样本后，460 份成功问卷样本中，已明确表态者有 400 份，占总有效问卷的 86.93%；而回答不知道或尚未决定或甚至拒绝回答者有 60 份，占总有效问卷的 13.07%。

设定总体模糊权数值

首先，打算经过对照分析，运用权数值理念探讨投票取向因素{政治表现、政治形象、候选政见、所属党派}的重要性程度。令排序量为第一重要者所对应的排序量值为 4；第二重要者给定排序量值为 3；第三重要者排序量值为 2；最不重要者的排序量值为 1；具有同等重要地位者，排序量值为所概括重要程度范围排序量值的平均。表 4.5 为 461 份所有成功问卷样本的实际回答状况。

表 4.5　访问成功样本的{政治表现、政治形象、候选政见、所属党派}重要程度排序量值出现次数

投票取向排序 / 因素	政治表现	政治形象	候选政见	所属党派
4	293	80	46	40
3.5	1	1	0	0
3	106	138	146	61
2.5	0	3	3	0
2	44	175	175	69
1.5	2	4	7	7
1	15	60	84	284

要使抽样样本分布接近总体特征结构，降低抽样误差，提升推断总体的精确度，令成功问卷样本经过反复加权法 $q = 2$ 反复加权后，如表 4.6 所示重新计算{政治表现、政治形象、候选政见、所属党派}的排序量值。

表 4.6　成功问卷样本反复加权后的{政治表现、政治形象、候选政见、所属党派}的重要程度排序量值出现次数

论域 / 排序值	政治表现	政治形象	候选政见	所属党派
4	294	77	41	45
3.5	1	1	0	0
3	103	142	145	57
2.5	0	5	5	0
2	45	171	188	63
1.5	1	3	6	6
1	16	61	75	289

$$R_1 = (4 \cdot 294) + (3.5 \cdot 1) + (3 \cdot 103) + (2.5 \cdot 0) + (2 \cdot 45) + (1.5 \cdot 1) + (1 \cdot 16) = 1\,596$$

$$R_2 = (4 \cdot 77) + (3.5 \cdot 1) + (3 \cdot 142) + (2.5 \cdot 5) + (2 \cdot 171) + (1.5 \cdot 3) + (1 \cdot 61) = 1\,157.5$$

$$R_3 = (4 \cdot 41) + (3.5 \cdot 0) + (3 \cdot 145) + (2.5 \cdot 5) + (2 \cdot 188) + (1.5 \cdot 6) + (1 \cdot 75) = 1\,071.5$$

$$R_4 = (4 \cdot 45) + (3.5 \cdot 0) + (3 \cdot 57) + (2.5 \cdot 0) + (2 \cdot 63) + (1.5 \cdot 6) + (1 \cdot 289) = 775$$

$$\sum_{i=1}^{4} R_i = 1\,596 + 1\,157.5 + 1\,071.5 + 775 = 4\,600$$

$$TW_1 = \frac{1\,596}{4\,600} = 0.35, TW_2 = 0.25, TW_3 = 0.23, TW_4 = 0.17$$

由此可得,{政治表现、政治形象、候选政见、所属党派} 的各自排序量总值为$R_1 > R_2 > R_3 > R_4$。

一般认为,影响投票决定的重要程度大小依序为政治表现、政治形象、候选政见、所属党派,也就是{政治表现、政治形象、候选政见、所属党派} 的总体模糊权重值为{0.35,0.25,0.23,0.17}。

设定个体模糊权数值

成功问卷样本经过反复加权法 $q = 2$ 反复加权后,分别计算个体模糊权数值 $SW_{i,j} = v_{i,j}/\sum_{i=1}^{i} v_{i,j}$, $v_{i,j}$ 为第j个受访者给予投票取向因素A_i的排序量值,$SW_{i,j}$为第j个受访者给予投票取向因素A_i的个体模糊权数值,其中$i = 1,2,3,4$, $j = 1,2,\cdots,461$。列举6个成功问卷样本的投票取向因素所属排序量值与个体模糊权数值如表4.7所示。

表 4.7 成功问卷样本的投票取向因素所属排序量值与个体模糊权数值

受访者	政治表现	政治形象	候选政见	所属党派	SW_1	SW_2	SW_3	SW_4
01	4	3	2	1	0.4	0.3	0.2	0.1
02	4	3	1	2	0.4	0.3	0.1	0.2
03	3	4	2	1	0.3	0.4	0.2	0.1
04	4	2	3	1	0.4	0.2	0.3	0.1
05	4	2	3	1	0.4	0.2	0.3	0.1
06	1	3	2	4	0.1	0.3	0.2	0.4

(a) 模糊权重设定：

设 $L_{i,j}$ 表示投票取向因素 A_i 的总体模糊权值与第 j 个受访者给予投票取向因素 A_i 的个体模糊权数值的乘积，即 $L_{i,j} = TW_i \times SW_{i,j}, i = 1,2,3,4, j = 1, 2,\cdots,461$，令 $W_{i,j}$ 为第 j 个受访者给予 A_i 的模糊权数值

$$W_{i,j} = L_{i,j} / \sum_{i=1}^{l} L_{i,j}$$

表 4.8 继续将 6 个成功问卷样本列举所属投票取向因素 A_i 的 $L_{i,j}$ 与模糊权数值 $W_{i,j}$ 如下。

表 4.8　成功问卷样本列举所属投票取向因素的模糊权数值

受访者	01	02	03	04	05	06
$L_{1,j}$	0.14	0.14	0.10	0.14	0.14	0.03
$L_{2,j}$	0.08	0.08	0.10	0.05	0.05	0.08
$L_{3,j}$	0.05	0.02	0.05	0.07	0.07	0.05
$L_{4,j}$	0.02	0.03	0.017	0.02	0.02	0.07
$W_{1,j}$	0.50	0.51	0.39	0.50	0.50	0.15
$W_{2,j}$	0.27	0.28	0.38	0.18	0.18	0.33
$W_{3,j}$	0.17	0.09	0.17	0.25	0.25	0.21
$W_{4,j}$	0.06	0.12	0.06	0.06	0.06	0.30

(b) 最大隶属度判定：

一旦提出模糊统计加权分类法后，将受访者针对投票取向因素形式，给予三位候选人的评定分数。经过模糊权重 $W_{i,j}$ 修正后，分别计算三位候选人在投票取向因素中所得的总评分数。

将{政治表现、政治形象、候选政见、所属党派}视为候选人属性，运用模糊理论的隶属度观念，期望不忽视且能够逼近受访者内心的模糊不确定性质。受访者的隶属函数，可由候选人属性相对所属的隶属度表示，即受访者对不同的候选人将有不尽相同的支持度，再以最大隶属度判定法则，预测受访者最终将会决定支持哪一位候选人。试继续将 6 个成功问卷样本中，受访者针对{政治表现、政治形象、候选政见、所属党派}给予三位候选人的评定分数由表 4.9 给出。经过模糊理论的模糊权重修正后，六位受访者给予各个候选人的支持隶属度由表 4.10 给出。

表 4.9 受访者对{政治表现、政治形象、候选政见、所属党派}的评定分数

论域 \ 受访者		1	2	3	4	5	6
政治表现	Damer	9	0	7	6	4	7
	Eden	7	10	6	9	7	7
	Ferry	6	6	0	5	9	7
政治形象	Damer	8	0	8	8	6	6
	Eden	9	10	6	7	8	6
	Ferry	7	7	0	9	9	5
候选政见	Damer	8	0	8	7	6	7
	Eden	8	10	6	8	6	6
	Ferry	7	8	0	5	7	7
所属党派	Damer	8	0	8	4	4	5
	Eden	6	10	6	6	6	7
	Ferry	4	6	0	4	8	7

表 4.10 经过模糊权重修正后,6 位受访者给予各个候选人支持隶属度

受访者		01	02	03	04	05	06
模糊权重修正论域的总评分	Damer	8.49	0	7.61	6.49	4.87	5.99
	Eden	7.64	10	6	8.56	6.86	6.38
	Ferry	6.31	6.45	0	5.66	8.42	6.25
给予候选人支持隶属度	Damer	0.38	0	0.56	0.31	0.24	0.32
	Eden	0.34	0.61	0.44	0.41	0.34	0.34
	Ferry	0.28	0.39	0	0.27	0.43	0.34
最大隶属度判定最终将支持的候选人		Damer	Eden	Damer	Eden	Ferry	Eden

即第一位受访者的隶属函数，可由候选人属性相对所属的隶属度表示为

$$\mu_A(x) = 0.38 I_{Damer}(x) + 0.34 I_{Eden}(x) + 0.28 I_{Ferry}(x)$$

再依据最大隶属度的判定原则，第一位受访者给予 Damer 的支持隶属度最高 $\max\{0.38, 0.34, 0.28\} = 0.38$。

经过整体考虑后，依据最大隶属度的判定原则，第一位受访者于三位候选人中，将决定支持 Damer。在 460 份成功问卷中，有 60 份受访者未决定投票意向的问卷。依据最大隶属度的判定原则，预测未决定投票意向的受访者，将支持 Damer 为 16 位，支持 Eden 为 30 位，支持 Ferry 为 14 位。

最大隶属度预测意向依据可靠性分析法尝试动态修正

进一步依据可靠性原则，将最大隶属度预测意向依据可靠性分析法，尝试动态修正意向预测，表 4.11 为已表态的受访者的投票意愿与最大隶属度判定的综合比较。

表 4.11　已表态受访者的投票意愿与最大隶属度判定的综合比较

隶属度判定 / 表态候选人	Damer	Eden	Ferry	总得票数
1. Damer	162	15	12	189
2. Eden	3	136	43	182
3. Ferry	0	2	27	29
总得票数	165	153	82	400

表 4.11 显示，在 400 位表态的受访者中，有 189 位表态支持 Damer，182 位支持 Eden，29 位支持 Ferry。但是，依据最大隶属度判定后，预测 165 位将支持 Damer，153 位支持 Eden，82 位支持 Ferry。因此，将最大隶属度预测意向依据可靠性分析法尝试动态修正。

在 400 位表态的受访者中，依据最大隶属度预测意向判定后，预测 165 位将支持 Damer 的受访者中，真正表态支持 Damer 的受访者有 162 位，故最大隶属度预测意向的可靠性动态修正为 162/165 = 0.98。未表态支持 Damer，却真正表态支持 Eden 的受访者有 3 位，因而最大隶属度预测意向的可靠性动态修正为 3/165 = 0.02。未表态支持 Damer，却真正表态支持 Ferry 的受访者有 0 位，所以最大隶属度预测意向的可靠性动态修正为 0。

在 400 位已表态的受访者中，依据最大隶属度预测意向判定后，预测 153 位将支持 Eden 的受访者中，真正已表态支持 Eden 的受访者共有 136 位，因此，最

大隶属度预测意向的可靠性动态修正为 136/153 = 0.89。

在 400 位表态的受访者中，依据最大隶属度预测意向判定后，预测 82 位将支持 Ferry 的受访者中，真正表态支持 Ferry 的受访者有 27 位，故最大隶属度预测意向之可靠性动态修正为 27/82 = 0.33。未表态支持 Ferry，却真正表态支持 Eden 的受访者有 43 位，因而最大隶属度预测意向的可靠性动态修正为 43/82 = 0.52；其余依此类推。依据可靠性分析法，我们尝试动态修正最大隶属度的意向预测，其原则表列举如表 4.12 所示。

表 4.12 依据可靠性分析修正最大隶属度判定的原则表

最大隶属度判定 / 表态的受访者	O_1	O_2	…	O_n	总得票数
O_1	O_{11}	O_{12}	…	O_{1n}	TO_1
O_2	O_{21}	O_{22}	…	O_{2n}	TO_2
⋮	⋮	⋮		⋮	⋮
O_n	O_{n1}	O_{n2}	…	O_{nn}	TO_n
总得票数	TMO_1	TMO_2	…	TMO_n	TO

把候选人已经明确表达的受访者中所获得的票数，与未明确表达的受访者中预测将获得的票数，作为加总处理来预测各位候选人的总得票数如下

Damer 得票数 $= 189 + 16 \cdot 0.982 + 30 \cdot 0.098 + 14 \cdot 0.146 = 210$

Eden 得票数 $= 182 + 16 \cdot 0.018 + 30 \cdot 0.889 + 14 \cdot 0.524 = 216$

Ferry 得票数 $= 29 + 16 \cdot 0 + 30 \cdot 0.013 + 14 \cdot 0.329 = 34$

为了更能显示各位候选人将获得选民的支持程度为何，试将预测候选人的总得票率列举如表 4.13 所示。

表 4.13 比较预测候选人得票率与实际候选人获得的得票率

	预测		实际	
候选人	得票数	得票率	得票数	得票率
1. Damer	210	46	688 072	45.91
2. Eden	216	47	766 377	51.13
3. Ferry	34	7	44 452	2.97
总　共	460	100	1 498 901	100

结果分析

近年来，民意调查已成为 ABC 市的新事物，民主社会强调凡事要以民意为

依归才能真正察觉人民所欲。ABC 市转型为大时代的趋势，1998 年执行第一届 ABC 市长选举，民调机构如火如荼地积极进行 ABC 市市长选情预测，所应用的数学理论预测模型不胜枚举，但尚未实际应用模糊理论预测选情。

本实例将政治表现，政治形象，候选政见，所属党派列为反映选民心态的影响因素，应用模糊理论以模糊统计分析传统问卷，来预测 ABC 市长选举情势。首先，就成功问卷的个人基本数据作样本代表性检验，为成功样本作一权重修正，以增加估计的精确度。本项研究认为，每一位选民将因自身独有的特征因素，而对影响投票取向四项因素有着不尽相同的重要程度，故考虑个体模糊权重与总体模糊权重的选定，将选民的模糊想法作一系统性分析。

我们分别侦测选民个人与全部选民判别影响投票取向四项因素的重要性程度，再将个体模糊权重与总体模糊权重的乘积设定为模糊权重，认为选民将会基于个人独有特质与外在大环境的影响，进而判别影响投票取向四项因素的重要性程度。本项研究请选民就候选人在{政治表现、政治形象、政见、所属党派}的投票取向因素上，其所能感受到的满意程度给予分数。此实例应用模糊论域，认为选民将依照自我的评分标准，给予候选人分数，即三位 ABC 市长候选人将各具有不同的支持度。然后，利用隶属度与决定投票意向之间的微妙关系，以最大隶属度预测选民最终决定的投票意向，并增加可靠性分析修正最大隶属度判定原则，提高最大隶属度判定的准确性，重新评估游离票，将对复杂的选举形势作更适合的预测估计。

本实例将成功样本反复加权后所获得的信息中，预测 Damer 得票数为 210，Eden 得票数为 216，Ferry 得票数为 34，即 Damer 得票率为 45%，Eden 得票率为 47%，Ferry 得票率为 7%。

1998年，ABC市长选举实际开票结果为，Damer得票数为 688 072，Eden得票数为 766 377，Ferry 得票数为 44 452，即 Damer 得票率为 45.91%，Eden 得票率为 51.13%，Ferry 得票率为 2.97%。不难发现，所预测 Damer 的得票率与实际开票的结果逼近，但是所预测 Eden 与 Ferry 之得票率与实际开票结果不一。

一些文献认为，在整个投票决定过程中的任何时机，都有可能让选民作出对谁投票的决定，所以将整个投票决定过程分成四个阶段分别观察：(1) 很早就决定了；(2) 竞选期间决定的；(3) 投票前一天才决定；(4) 投票当天决定。事实上，在选举最后一刻，候选人及政党的竞选行为，确实会对竞选结果产生影响，这在竞选活动密集于法定期间的环境中，是预测选举困难的原因之一。一些文献认为，选民可能是候选人最忠实的拥护和一贯的支持者；或看到候选人的宣传或竞选活动，受到吸引或认同；可能是受到请托或组织动员；可能在公办或

私办政见发表会上，对某位候选人留有特别的印象。因此，个人的投票结果仍需参考动态的宣传与竞选活动效果，比如宣传口号、贿选、政见发表会等等。

本研究民意调查进行期间为1998年11月23日至12月3日，而1998年ABC市长选举于1998年12月5日举行。在选举期间，由于Eden与Damer的民意支持度相当接近，民主党与共和党皆认为选举形势告急，无不在文化宣传战，组织战，省籍情结，候选人个人特质等等，加倍努力大作文章。共和党为巩固票数与开拓票源，费尽不少心思在宣传保卫战上，努力经营选民情绪，巩固支持Eden的选票，企图拉拢徘徊在Eden与Ferry间犹豫不决的选民，甚而鼓吹支持Ferry的选民，将决定性的一票投给最有希望当选ABC市长的Eden。本研究分析过程中亦发现，选民虽然认为Ferry整体感觉优于Eden，但却决意将票投给Eden的比率是很高的。

修正最大隶属度判定的原则中，虽然最大隶属度判给Ferry，但是表态投给Eden的受访者，占最大隶属度判给Ferry的总受访者的52.4%，或许可以看出一些端倪。

在传统问卷的设计形式上，受限于只能让受访者在数个选项中勾选出一个答案，即受访者将经由内心反复地评估，将依照对事态的平均看法来回答，在0到10的标尺上填选一个较适当的标尺分数位置。然而，问卷题目所表达的语言，也因为受访者个人差异性的存在，会造成对衡量标尺的认知程度不同，即不同的受访者或许会对标尺上7的位置给予不尽同等的看法。并且，个人对标尺上8的位置未必相较于另一受访者对标尺上6的位置代表较高的认同度。

然而，人类语言的本质原来就具备模糊性，其语言表达方式隐藏着不确定性，传统问卷很难真实地显现出受访者心态的分布状况。或许未来研究者可以以设计模糊问卷为努力的方向，改革传统问卷上不可避免的缺陷。模糊问卷的设计形式，允许受访者按心中的感受程度于不同的选项上填写反映状态的百分比，或请受访者于语意尺度上填写评分范围等等回答项设计方式，想必更能真切地反映受访者对题目感受的分布形势，也包容了语言本质的模糊性。然而，在进行电话民意调查的过程中，必将面临受访者有填答问卷的困扰，如何将题意正确地传达给受访者，并告知受访者依照内心感受分布程度填答问卷，势必成为电话访问模糊问卷迫切解决之处。倘若考虑应用模糊问卷进行民意面访调查，人力与时间的高成本付出方式，将成为考虑可行性的阻挠原因之一。

选情民意调查问卷除了将政治表现、政治形象、候选政见、所属党派列为反映选民心态的影响因素外，进一步再考虑加入其他政治影响变量，并且因应当时的选举形势加入政治策略性的相关题目，那么预测选情的问卷题目将更趋完

备。假若能圆满地解决受访者填答之困难处与付出成本的考虑，那么应用模糊统计分析模糊问卷，将会引导民意调查预测于新的领域。

4.5　结　论

自从以模糊集合的概念作为测试不明确隶属度的工具以来，许多的模糊研究就尝试以此方法设定理论框架，并尝试以此理论作为建构更为先进模糊理论的基础。模糊理论在社会科学中的应用迄今虽取得一些进展，但是，在统计学中的应用还是刚刚兴起。

虽然在心理计量学有一些模糊测量表的设计，作为测量态度的依据，以及某些态度与偏好，比如社会学中对于理想子女数以及男女教育年限的应用及测量(Li,1988)。但是，大部分的研究对象均局限于大学主修社会科学的学生的样本中，这些研究几乎不曾将扎德所发展隶属度的概念在一般人口样本中加以测试。

本章研究的重要贡献在于，利用"模糊集合"概念所发展出来的隶属度概念应用于一般大型实地调查的样本中，并且尝试以模糊逻辑所发展出来的模糊测量量表，更进一步地用模糊统计作为分析数据的依据。

在社会科学研究的测量理念上，模糊统计和模糊相关性的使用可说是数字模式的推广。如果利用假定的"精确值"，那么就可能导致错误地建立模型，也可能扩大研究结果和实际状态之间的误差。因此，具有模糊性的数值数据，在计量方法的演算过程中，是否符合人类的逻辑推论与归纳的原则，是值得人们进一步思考的专题。

近年来，在人文社会科学的测度理念里，模糊统计和模糊相关性日渐受到重视，这应是复杂的人文社会科学现象无法以传统数值模型充分合理解释的一种自然发展结果。曼斯基(Manski,1990)曾经指出，数字数据有需求过度及过度解释的危险，如果应用模糊的数值数据，就能避免这样的危险发生。有鉴于此，本章尝试应用区间值的模糊特性所作的分析调查，可使研究者在处理不确定情况时能较符合实际状况。

最后，我们认为，发展模糊语言系统与计算机人工智能计算，共同应用于思维函数的估计研究是一个新颖的交叉方向，而且已经取得一些成果。此研究方向将会为模糊数据统计提供有力的工具。另外，为了得到合适的感觉准确度，软计算必将成为值得继续深入研究探讨的技术，且在社会科学领域，激励更多未来模糊数据统计调查方法的发展。

附表 4.1　乐观测量表

请就符合你的状况回答。回答时请用分布方式作答，在每一个陈述句后面，选择符合你状况的选项，可以是一种、两种、三种、四种或五种，再以"1"作分布，填人适合比例分布，请注意，每一题的分布加起来都要是"1"。例如，陈述句为"我是健康的"，您若觉得，你有时符合"不同意"，但有时符合"非常同意"，则您可能的填法如下：

	非常不同意	不同意	普通	同意	非常同意
我是健康的		0.3			0.7

乐观问项	非常不同意	不同意	普通	同意	非常同意
1.在不确定的时刻，我通常是作最好的打算。					
2.如果有什么事情可对我不利，其结果真的对我不利。					
3.看事情，我总是看光明的一面。					
4.对我自己的前途，我总是乐观的。					
5.我几乎从不期望事情如我所愿。					
6.事情发展的结果，从来就没有符合我的需求。					
7.我相信否极泰来，塞翁失马，焉知非福的观念。					
8.我极少期望好事会降临在我身上。					

附表4.2　ABC市长选民投票意向问卷

您好，我们正在作有关ABC市长选民投票意向为主题的研究，可不可以耽误您一些时间请您回答以下几个问题，谢谢！

1.就Damer、Eden、Feery三位ABC市长候选人的政治表现来看，现在请您评分，假设评分的标准为0至10,0分表示很不满意，10分表示很满意，5分表示普通：

(1) 请问您会给Damer的政治表现几分？________

(2) 请问您会给Eden的政治表现几分？________

(3) 请问您会给Feery的政治表现几分？________

2.就Damer、Eden、Feery三位ABC市长候选人各自所建立的政治形象来看

(1) 请问您会给Damer所建立的政治形象几分？________

(2) 请问您会给Eden所建立的政治形象几分？________

(3) 请问您会给Feery所建立的政治形象几分？________

3.就Damer、Eden、Feery三位ABC市长候选人各自发表的政见来看

(1) 请问您会给Damer所发表的政见几分？________

(2) 请问您会给Eden所发表的政见几分？________

(3) 请问您会给Feery所发表的政见几分？________

4.就Damer、Eden、Feery三位ABC市长候选人各自所属的党派，给予您的印象来看

请问您会给共和党几分？________。给民主党几分？________。给其他小党几分？________。

5.影响您的投票决定的下面四个因素中：

1.候选人的政治表现；2.竞选所发表的政见；3.所建立的政治形象；4.所属党派

请问哪一个因素最重要________；第二重要________；第三重要________；第四重要________。

6.如果明天就是ABC市长选举投票日，请问您会不会去投票？

会(1)________；不会(2)________；不知道(3)________。

7.如果明天就是ABC市长选举投票日，就Damer、Eden、Feery三位ABC市长候选人，请问您的选票最有可能投给哪位候选人？

(1)Damer________　(2)Eden________　(3)Feery________

(4) 不知道________

个人资料:

请问您是那一年出生?________ 性别:男________、女________

请问您的最高学历是什么:小学、中学、高中职、大专、大学、硕士、博士。

省籍:请问您的父亲是本省人________、其他省人________、其他国移民________。

请问您的户籍是在 ABC 市________区。

我们的访问就到此结束,谢谢你接受我们的访问。

第 5 章　模糊数据均数估计

在传统的统计推断方法中，为了了解未知总体参数值，我们经常借助于一些评估准则，找出适当的统计量来对总体参数进行估计。平均数是认识总体集中趋势最重要的总体参数之一，人们常常利用无偏估计量，比如样本平均数来估计总体集中趋势。然而，在日常生活中，总体平均数经常是带有模糊、不确定性的语意变量，或是一个可能区间，传统的估计量评估准则及估计方法便无法适用于此种情况。

本章尝试以软计算方法，结合模糊理论，提出对模糊数据总体平均数进行估计的方法。同时，针对模糊参数的估计量，我们提出适当可行的评判准则，希望借此能找出模糊数据参数的最佳估计量。最后，我们把这种最佳估计量应用于实证数据。

5.1　模糊数据总体均数

假定随机变量 X 的总体概率密度函数为 $f(x \mid \theta)$，其中 θ 表示未知参数。为了估计未知参数 θ，我们从总体中随机抽取一组样本，得到观测值为 $x_1, x_2, \cdots, x_n$。利用点估计方法计算出一个估计量(Estimator)，用 $\hat{\theta}$ 表示。再把观测值 $x_1, x_2, \cdots, x_n$ 代入估计量中得到一个数值，称此数值为参数 θ 的估计值。

点估计就是指从总体抽取随机样本，经由样本统计量来估计总体的参数。比较常用的估计方法有(1) 最大似然法(Maximum Likelihood Method，又称极大似然法)；(2) 矩方法(Moment Method)。

【定义 5.1】　离散型模糊总体均数(Fuzzy Population Mean for Discrete Type)

设 U 表示一个论域，令 $L = \{L_1, L_2, \cdots, L_k\}$ 为论域 U 上的 k 个语言变量，$\left\{x_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}, i = 1,2,\cdots,n\right\}$为模糊数所集合成的总体。把离散型模糊总体均数定义为

$$F\mu = \frac{\frac{1}{N}\sum_{i=1}^{N} m_{i1}}{L_1} + \frac{\frac{1}{N}\sum_{i=1}^{N} m_{i2}}{L_2} + \cdots + \frac{\frac{1}{N}\sum_{i=1}^{N} m_{ik}}{L_k} \tag{5.1}$$

记为 $F\mu$，其中 m_{ij} 表示第 i 个模糊数相对于 L_j 的隶属度，$i = 1,2,\cdots,N, j = 1,2,\cdots,k$。

【定义 5.2】 连续型模糊总体均数(Fuzzy Population Mean for Continuous Type)

设 U 表示一个论域，令 $\{x_i = [a_i, b_i], i = 1,2,\cdots,N\}$ 为论域 U 里的模糊区间所构成的总体，则把连续型模糊总体均数定义为

$$F\mu = \left[\frac{1}{N}\sum_{i=1}^{N} a_i, \frac{1}{N}\sum_{i=1}^{N} b_i\right] \tag{5.2}$$

记为 $F\mu$。

【定义 5.3】 两个离散型模糊样本相对距离

设 U 表示一个论域，令 $L = \{L_1, L_2, \cdots, L_k\}$ 为论域 U 上的 k 个语言变量，而$\left\{x_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}, i = 1,2,\right\}$，$\sum_{j=1}^{k} m_{ij} = 1$ 为从论域 U 中抽出的两个模糊样本。则两个离散型模糊样本 x_1 与 x_2 的相对距离定义为

$$\delta(x_1, x_2) = \frac{|m_{11} - m_{21}|}{L_1} + \frac{|m_{12} - m_{22}|}{L_2} + \cdots + \frac{|m_{1k} - m_{2k}|}{L_k} \tag{5.3}$$

可用式(5.3)表示两模糊数的相似程度。若 $L = \{L_1, L_2, \cdots, L_k\}$ 为一组有序尺度，例如，$L_1 = 1, L_2 = 2, \cdots, L_5 = 5$，则我们可以把离散型模糊样本的距离，又称为绝对距离，定义如下。

【定义 5.4】 两个离散型模糊样本的距离

设 U 表示一个论域，令 $L = \{L_1 = 1, L_2 = 2, \cdots, L_k = k\}$ 为论域 U 上的 k 个一组有序语言变量。若$\left\{x_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}, i = 1,2\right\}$，$\sum_{j=1}^{k} m_{ij} = 1$ 为从论域 U 中抽出的两个模糊样本。把两个离散型模糊样本 x_1 与 x_2 的距离定义为

$$d(x_1, x_2) = \frac{1}{k-1}\left|\sum_{j=1}^{k}(m_{1j} - m_{2j}) \cdot j\right| \tag{5.4}$$

式(5.4)中分母的 k 表示有序语言变量的总个数，$i = 1,2,\cdots,N, j = 1,2,\cdots,k$。

由定义可知，距离 d 最大为 1，最小为 0。d 值愈小，表示两个模糊样本愈近似；d 值愈大，表示两个模糊样本愈不近似。值得注意的是，此距离的定义不符

合传递性(Transitivity)，但比较符合我们日常生活的实际经验。若"A 与 B 的看法差不多"，"B 与 C 的看法差不多"，不见得就表示"A 与 C 的看法差不多"，因为 A，C 的想法可能同中有异。从下面例 5.1 可以看出，式(5.4) 的实际含义。

【例 5.1】　若一组受访者的模糊意见表示如表 5.1 所示。

表 5.1　受访者的模糊意见

您对执政党 2005 年施政满意度	1 = 很不满意	2 = 不满意	3 = 普通	4 = 满意	5 = 很满意
x_1	0	0	0	0	1
x_2	1	0	0	0	0
x_3	0	0	1	0	0
x_4	0	0.5	0	0.5	0
x_5	0	0	0.5	0	0.5
x_6	0	0	0.8	0	0.2

根据离散型模糊样本相对距离定义，我们可算出各样本间的相对距离。从相对距离的模糊数值大小可看出，各个样本在不同语意变量喜好的差异性，例如

$$\delta(x_1,x_2)=\frac{|0-1|}{L_1}+\frac{|0-0|}{L_2}+\frac{|0-0|}{L_3}+\frac{|0-0|}{L_4}+\frac{|1-0|}{L_5}=$$

$$\frac{1}{L_1}+\frac{0}{L_2}+\frac{0}{L_3}+\frac{0}{L_4}+\frac{1}{L_5}$$

$$\delta(x_5,x_6)=\frac{|0-0|}{L_1}+\frac{|0-0|}{L_2}+\frac{|0-0|}{L_3}+\frac{|0.8-0.5|}{L_4}+\frac{|0.2-0.5|}{L_5}=$$

$$\frac{0}{L_1}+\frac{0}{L_2}+\frac{0}{L_3}+\frac{0.3}{L_4}+\frac{0.3}{L_5}$$

$\delta(x_1,x_2)$ 的模糊数值显示，x_1 与 x_2 在非常不满意与非常满意这两个选项的差异性很大，而其他选项的差异性则无。$\delta(x_5,x_6)$ 的模糊数值则显示 x_5 与 x_6 在普通和很满意这两个选项的差异性，而其他选项的差异性则无。

而在语意变量为顺序尺度情况下，由定义 5.4，我们可进一步算出各个样本之间的(绝对) 距离值。从绝对距离的数值大小可以看出，各个样本整体喜好程度的差异性。根据离散型模糊样本绝对距离定义，可以分别计算 $d(x_i,x_j)$，比如下表 5.2。

表 5.2 受访者的模糊意见距离

$d(x_i,x_j)$	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0	1	0.5	0.5	0.75	0.4
x_2		0	0.5	0.5	0.25	0.6
x_3			0	0	0.25	0.1
x_4				0	0.5	0.1
x_5					0	0.48
x_6						0

由上表我们发现，即 x_3 受访者的意见与 x_4 受访者的意见最接近。而 x_3 受访者的意见与 x_6 受访者的意见以及 x_4 受访者的意见与 x_6 受访者的意见次之；x_1 受访者与 x_2 受访者的意见最不接近。

【定义 5.5】 连续型模糊样本的距离

设 U 表示一个论域，令 $\{x_i=[a_i,b_i],i=1,2\}$ 为从论域 U 中抽出的两个模糊样本，$c_i=\dfrac{a_i+b_i}{2}$，$l_i=b_i-a_i$。把两个连续型模糊样本 x_1 与 x_2 的距离定义为

$$d(x_1,x_2)=|c_1-c_2|+(1-e^{-|l_1-l_2|}) \tag{5.5}$$

用此式表示两模糊样本的相似程度，d 值愈小，表示两模糊样本愈近似；d 值愈大，表示两模糊样本愈不近似。

【例5.2】 哈尔滨某大学2006年调查5位不同学科的硕士毕业生的预期年薪，结果如表 5.3 所示。

表 5.3 硕士毕业生的预期年薪

毕业生	预期年薪(万元)	c	l
A	[2,3]	2.5	1
B	[3,4]	3.5	1
C	[3,5]	4	2
D	[2,4]	3	2
E	[3,7]	5	4

根据连续型模糊样本绝对距离定义 5.5，可得 5 位毕业生预期年薪的意见距离，如表 5.4 所示。

表 5.4　5 位硕士毕业生预期年薪的意见距离

d	A	B	C	D	E
A	0	1	2.13	1.13	3.45
B		0	1.13	1.13	2.45
C			0	1	1.86
D				0	2.86
E					0

由上表可以发现，即 A，B 受访者意见，C，D 受访者意见最接近。而 A 受访者的意见与 E 受访者的意见最不接近。

对于一组数据 $X_1, X_2, \cdots, X_n$ 来说，我们想看其分散程度通常是求样本方差 $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$。但是，对于模糊样本来说，其数据形式为多值或是区间，我们该如何分析这种模糊样本数据的离散性(Diffusibility)呢？

注意到，对于离散型模糊样本来说，考虑用相对于每一语言变量(或分类尺度)计算其方差。然而，由于区间本身就具有离散性，长度越长则离散程度越大，因此我们对连续型模糊样本不仅要考虑中心点离散度，还需考虑区间长度的分散(Spread)程度。

【定义 5.6】　离散型模糊方差(Fuzzy Variance for Discrete Type)

设 U 表示一个论域，令 $L = \{L_1, L_2, \cdots, L_k\}$ 为论域 U 上的 k 个有序变量，而 $\left\{x_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}, i = 1,2,\cdots,N\right\}$，$\sum_{j=1}^{k} m_{ij} = 1$ 为论域中的总体观测值，其总体均数

$$F\mu = \frac{\mu_1}{L_1} + \frac{\mu_2}{L_2} + \cdots + \frac{\mu_k}{L_k}$$

$$\left\{x_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}, i = 1,2,\cdots,n\right\}, \sum_{j=1}^{k} m_{ij} = 1$$

为从论域中抽出的一组样本观测值，其样本均数为

$$F\bar{X} = \frac{\hat{\mu}_1}{L_1} + \frac{\hat{\mu}_2}{L_2} + \cdots + \frac{\hat{\mu}_k}{L_k}$$

把离散型模糊总体方差(Fuzzy Population Variance for Discrete Type)定义为

$$F\sigma^2 = \frac{\frac{1}{N}\sum_{i=1}^{N}(m_{i1}-\mu_1)^2}{L_1} + \frac{\frac{1}{N}\sum_{i=1}^{N}(m_{i2}-\mu_2)^2}{L_2} + \cdots + \frac{\frac{1}{N}\sum_{i=1}^{N}(m_{ik}-\mu_k)^2}{L_k} \tag{5.6}$$

而把离散型模糊样本方差(Fuzzy Sample Variance for Discrete Type)定义为

$$FS^2 = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(m_{i1}-\hat{\mu}_1)^2}{L_1} + \frac{\frac{1}{n-1}\sum_{i=1}^{n}(m_{i2}-\hat{\mu}_2)^2}{L_2} + \cdots + \frac{\frac{1}{n-1}\sum_{i=1}^{n}(m_{ik}-\hat{\mu}_k)^2}{L_k} \tag{5.7}$$

【定义 5.7】 连续型模糊方差(Fuzzy Variance for Continuous Type)

设 U 表示一个论域,令 $\{x_i=[a_i,b_i], i=1,2,\cdots,N\}$ 为论域 U 里的模糊区间所构成的总体,其总体均数为 $F\mu=[\mu_l,\mu_u]$;$\{x_i=[a_i,b_i], i=1,2,\cdots,n\}$ 为从论域 U 里抽出的一组随机模糊区间样本,其样本均数为 $F\overline{X}=[\hat{\mu}_l,\hat{\mu}_u]$。令 $c_i=\frac{a_i+b_i}{2}$,$l_i=|b_i-a_i|$,$c=\frac{\mu_l+\mu_u}{2}$,$l=|\mu_u-\mu_l|$,$\hat{c}=\frac{\hat{\mu}_l+\hat{\mu}_u}{2}$,$\hat{l}=|\hat{\mu}_u-\hat{\mu}_l|$。把连续型模糊总体方差(Fuzzy Population Variance for Continuous Type)定义为

$$F\sigma^2 = \left< \frac{\sum_{i=1}^{N}(c_i-c)^2}{N}, \frac{\sum_{i=1}^{N}(l_i-l)^2}{N} \right> \tag{5.8}$$

而把连续型模糊样本方差(Fuzzy Sample Variance for Continuous Type)定义为

$$FS^2 = \left< \frac{\sum_{i=1}^{n}(c_i-\hat{c})^2}{n-1}, \frac{\sum_{i=1}^{n}(l_i-\hat{l})^2}{n-1} \right> \tag{5.9}$$

式(5.8),(5.9)右边第一项表示总体区间中心的离散程度,第二项表示总体(样本)区间长度的离差程度。

【例 5.3】 若有 A,B,C 三位受访者的意见可表示为模糊数形式如下

$$X_A = \frac{0}{L_1}+\frac{0}{L_2}+\frac{0.5}{L_3}+\frac{0.5}{L_4}+\frac{0}{L_5}$$

$$X_B = \frac{0}{L_1}+\frac{0.2}{L_2}+\frac{0.8}{L_3}+\frac{0}{L_4}+\frac{0}{L_5}$$

$$X_C = \frac{0}{L_1}+\frac{0}{L_2}+\frac{0.7}{L_3}+\frac{0.3}{L_4}+\frac{0}{L_5}$$

则其模糊样本方差为

$$FS^2 = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(m_{i1}-\hat{\mu}_1)^2}{L_1} + \frac{\frac{1}{n-1}\sum_{i=1}^{n}(m_{i2}-\hat{\mu}_2)^2}{L_2} + \cdots + \frac{\frac{1}{n-1}\sum_{i=1}^{n}(m_{ik}-\hat{\mu}_k)^2}{L_k} =$$

$$\frac{0}{L_1} + \frac{0.013}{L_2} + \frac{0.023}{L_3} + \frac{0.063}{L_4} + \frac{0}{L_5}$$

此模糊样本方差所代表的意义为：很不满意的隶属度为 0，不满意的隶属度为 0.01，无意见的隶属度为0.02，满意的隶属度为0.06，很满意的隶属度为0。此模糊方差是一个模糊数，它表现出此访问调查中，受访者对于满意的程度歧异性较大，而对于很不满意、很满意大家是一致地没有这方面的感受。也就是，我们可利用模糊方差相对于各语意变量的隶属度大小，得知模糊样本值的一致性。

【例 5.4】　表 5.5 为 5 位商家对店面面宽需求的隶属度选择，想要了解此次调查结果的可靠性。

表 5.5　5 位受访者对店面面宽需求的隶属度选择

商　家	店面面宽	c	l
A	[2,5]	3.5	3
B	[6,10]	8	4
C	[7,12]	9.5	5
D	[3,4]	3.5	1
E	[17,20]	18.5	3

计算模糊样本方差如下

$$FS^2 = < \frac{\sum_{i=1}^{n}(c_i-\hat{c})^2}{n-1}, \frac{\sum_{i=1}^{n}(l_i-\hat{l})^2}{n-1} > = < 38.4, 2.3 >$$

5.2　模糊数据总体均数最佳估计方法

对于模糊总体平均数的估计方法来说，除了前述的模糊样本均数外，下面我们给出几个可能的最佳估计方法，并提出可行的估计量评估准则，以此比较各个估计量的优劣。

模糊最大隶属函数估计法

传统的统计学推断方法中，经常通过求最大似函数的最大值来获得未知总

体参数的最大似然估计量。类似地，在模糊统计分析中，我们尝试通过求隶属函数的最大值来获得模糊总体均数的估计量。

【定义 5.8】 离散型模糊最大隶属度估计量

设 U 表示一个论域，令 $L=\{L_1,L_2,\cdots,L_k\}$ 为论域 U 上的 k 个有序变量，而 $\left\{x_i=\frac{m_{i1}}{L_1}+\frac{m_{i2}}{L_2}+\cdots+\frac{m_{ik}}{L_k}, i=1,2,\cdots,n\right\}$，$\sum_{j=1}^{k} m_{ij}=1$ 为从论域中抽出的一组模糊样本。

令 $S_j=\frac{1}{n}\sum_{i=1}^{n} m_{ij}\cdot I_{ij}$，其中 $I_{ij}=\begin{cases}1, 若\ m_{ij}=\max(m_{i1},\cdots,m_{ik})\\0, 若\ m_{ij}\neq\max(m_{i1},\cdots,m_{ik})\end{cases}$，$i=1,2,\cdots,N$，$j=1,2,\cdots,k$。把离散型模糊总体均数 $F\mu$ 的模糊最大隶属度估计量（Fuzzy Maximum Membership Estimator, MM, for Fuzzy Mean with Discrete Type $F\mu$）定义为

$$F_{\hat{\mu}_{MM}}=\frac{\frac{S_1}{\sum S_j}}{L_1}+\frac{\frac{S_2}{\sum S_j}}{L_2}+\cdots+\frac{\frac{S_k}{\sum S_j}}{L_k} \tag{5.10}$$

【例 5.5】 离散型最大隶属度估计量应用于教学满意度调查。

针对哈尔滨市某高校想了解学生对某老师教学方式的满意程度进行调查，若将满意程度分为：L_1 = 很不满意，L_2 = 不满意，L_3 = 普通，L_4 = 满意，L_5 = 很满意等五种，学校以此打算了解学生对某老师教学方式的满意程度。随机抽取某个教学班级 6 位学生作调查，每位学生对此老师的满意程度隶属度如表 5.6 所示。

表 5.6 6 位受访学生对某老师教学满意度的隶属度选择

学生	很不满意 L_1	不满意 L_2	普通 L_3	满意 L_4	很满意 L_5
1	0	0	0	1*	0
2	0	0	0.3	0.7*	0
3	0	0.8*	0.2	0	0
4	0	0	0	0.5*	0.5*
5	0	0	0.4	0.6*	0
6	1*	0	0	0	0

注：用上标“*”记一行中最大的数。

将表 5.6 的各位学生对满意度 5 个选项中筛选最大隶属度者留下，而选项中非最大隶属度者则给予隶属度 0，可得表 5.7 如下。

表5.7　6位受访学生对某老师教学满意度的最大隶属度

学生	很不满意	不满意	普通	满意	很满意
1	0	0	0	1	0
2	0	0	0	0.7	0
3	0	0.8	0	0	0
4	0	0	0	0.5	0.5
5	0	0	0	0.6	0
6	1	0	0	0	0
S_j	0.17	0.13	0	0.47	0.08
$S_j/\sum S_j$	0.20	0.16	0	0.55	0.09

其中

$$S_1 = \frac{0+0+0+0+0+1}{6} \approx 0.17$$

$$S_2 = \frac{0+0+0.8+0+0+0}{6} \approx 0.13$$

$$S_3 = \frac{0+0+0+0+0+0}{6} = 0$$

$$S_4 = \frac{1+0.7+0+0.5+0.6+0}{6} \approx 0.47$$

$$S_5 = \frac{0+0+0+0.5+0+0}{6} \approx 0.08$$

由上表可得，模糊总体平均数的模糊最大隶属度估计量为

$$F_{\hat{\mu}_{MM}} = \frac{0.17/0.85}{L_1} + \frac{0.13/0.85}{L_2} + \frac{0}{L_3} + \frac{0.47/0.85}{L_4} + \frac{0.08/0.85}{L_5} =$$

$$\frac{0.20}{L_1} + \frac{0.16}{L_2} + \frac{0}{L_3} + \frac{0.55}{L_4} + \frac{0.1}{L_5}$$

其代表的含义为：学生对此教师的满意程度，很不满意的隶属度大约为0.2，不满意的隶属度约有0.16，满意的隶属度大约为0.55，很满意的隶属度大约为0.1。由此例可看出，模糊总体平均数的模糊最大隶属度估计量不容易受离群值影响，例如 X_6，至于离散型模糊样本的离群值如何定义，则不在本书的探讨范围。

【定义5.9】　连续型模糊最大隶属度估计量

设 U 表示一个论域，令 $\{X_i = [a_i, b_i], i = 1, 2, \cdots, n\}$ 为从 U 中抽出的一

组模糊区间样本，而对应的模糊区间的长度为 $l_i = b_i - a_i$，相应的模糊数中心为 $CF_i = \frac{1}{2}(a_i + b_i)$。对任意的 j，令

$$M_j = \frac{\sum_{i=1}^{n} I_{[a_i, b_i]}(CF_j)}{n} \tag{5.11}$$

表示 CF_j 相对于 n 个模糊区间数的隶属度。

若 $M_r = \max(M_i), i = 1,2,\cdots,n$ 对应于模糊数中心 CF_r，把连续型模糊总体平均数 $F\mu = [a, b]$ 的模糊最大隶属度估计量(Fuzzy Maximum Membership Estimator, MM, for Fuzzy Mean with Continuous Type $F\mu$)定义为

$$F\hat{\mu}_{MM} = [\hat{a}_{MM}, \hat{b}_{MM}] =$$

$$\left[median(CF_r) - \frac{1}{2} median(l_i), median(CF_r) + \frac{1}{2} median(l_i)\right]$$

其中，$median(CF_\gamma)$ 表示具有最大隶属度的所有 CF_r 的中位数，$median(CF_\gamma)$ 表示所有 l_i 的中位数，$i = 1,2,\cdots,n$。

【例 5.6】 连续型模糊最大隶属度估计量应用于等候时间调查。

对快餐店业者来说，缩短顾客等候时间是非常重要的，不但可提升顾客对服务的满意度，也可以增加销售金额。哈尔滨市某快餐店业者采用开放式问卷调查，随机抽取 6 位顾客并调查他们可忍受的最长等候时间区段，结果如表5.8所示。

表 5.8 6 位顾客对等候时间的可忍受时间区段(分钟)

顾客	下限	上限	CF_i	M_i	l_i
1	4	6	5	0.5*	2
2	7	10	8.5	0.17	3
3	2	5	3.5	0.5*	3
4	2	8	5	0.5*	6
5	1	4	2.5	0.5*	3
6	1	2	1.5	0.33	1
$F\bar{X}$	2.83	5.83			

其中各样本 X_i 的模糊数中心 CF_i 相对于整体数据的隶属度 M_i 的计算如下

$$M_1 = \frac{\sum_{i=1}^{n} I_{[a_i,b_i]}(CF_1)}{n} =$$

$$\frac{I_{[4,6]}(5) + I_{[7,10]}(5) + I_{[2,5]}(5) + I_{[2,8]}(5) + I_{[1,4]}(5) + I_{[1,2]}(5)}{6} =$$

$$\frac{3}{6} = 0.5$$

同理,可得

$$M_2 = \frac{\sum_{i=1}^{n} I_{[a_i,b_i]}(CF_2)}{n} = \frac{1}{6} \approx 0.17$$

$$M_3 = \frac{\sum_{i=1}^{n} I_{[a_i,b_i]}(CF_3)}{n} = \frac{3}{6} = 0.5$$

$$M_4 = \frac{\sum_{i=1}^{n} I_{[a_i,b_i]}(CF_4)}{n} = \frac{3}{6} = 0.5$$

$$M_5 = \frac{\sum_{i=1}^{n} I_{[a_i,b_i]}(CF_5)}{n} = \frac{3}{6} = 0.5$$

$$M_6 = \frac{\sum_{i=1}^{n} I_{[a_i,b_i]}(CF_6)}{n} = \frac{2}{6} \approx 0.33$$

M_i 中最大隶属度为 $0.5 = M_1 = M_3 = M_4 = M_5$,而

$$median(CF_\gamma) = median(CF_1, CF_3, CF_4, CF_5) = median(5,3.5,5,2.5) = 4.25$$

$$median(l_i) = median(2,3,3,6,3,1) = 3$$

故 $F\mu$ 的模糊最大隶属度估计量为

$$F\hat{\mu}_{MM} = \left[4.25 - \frac{3}{2}, 4.25 + \frac{3}{2}\right] = [2.75, 5.75]$$

此估计量表示的含义为,顾客可忍受的最长等候时间区段最可能在 2.75 ~ 5.75 min 之间,因此,若该店店长观察顾客等候时间超过这个区间,可能就要采取适当的策略以缩短顾客等候时间,从而提升顾客满意度及增加营业销售金额。同样地,连续型模糊样本平均数也不容易受到样本中离群值的影响。例如,$x_2 = [7,10]$ 范围与大部样本几乎没有交集;$x_4 = [2,8]$ 虽然涵盖了可能的样本平均数区间,但所取的区间又太长。而采用连续型模糊最大隶属度估计量来估计 $F\mu$,一样也可以避免受到样本中离群值的影响。

5.3 模糊数据估计量的评判准则

无偏估计量(Unbiased Estimator)成为评估一个统计量是否合适与准确的重要准则。本节考察以模糊无偏估计量来评估模糊估计值的效率。在评估前，我们首先给出有关模糊期望值的几个定义。

【定义 5.10】 离散型总体的模糊期望值

设 $X=\frac{X_1}{L_1}+\frac{X_2}{L_2}+\cdots+\frac{X_k}{L_k}$ 表示总体中的一个模糊随机变量，其中 X_i 表示相对于 L_i 的随机变量，$0\leqslant E(X_i)\leqslant 1$，且 $\sum_{i=1}^{k}E(X_i)=1$。把离散型总体的模糊期望值定义为

$$FE(X)=\frac{EX_1}{L_1}+\frac{EX_2}{L_2}+\cdots+\frac{EX_k}{L_k}$$

【定义 5.11】 连续型总体的模糊期望值

设 $X=[X_l,X_u]$ 表示总体中的一个模糊随机变量，其中 X_l,X_u 表示相对于区间左、右端点的随机变量，且 $E(X_l)\leqslant E(X_u)$。把连续型总体的模糊期望值定义为

$$FE(X)=[E(X_l),E(X_u)]$$

【定义 5.12】 离散型总体参数的模糊无偏估计量

设 $F\hat{\theta}=\frac{\hat{\theta}_1}{L_1}+\frac{\hat{\theta}_2}{L_2}+\cdots+\frac{\hat{\theta}_k}{L_k}$ 表示一个离散型模糊总体参数 $F\theta=\frac{m_1}{L_1}+\frac{m_2}{L_2}+\cdots+\frac{m_k}{L_k}$ 的估计量，若 $E(F\hat{\theta})=F\theta$，则称 $F\hat{\theta}$ 为离散型模糊总体参数 $F\theta$ 的模糊无偏估计量。

【定义 5.13】 连续型总体参数的模糊无偏估计量

设 $F\hat{\theta}=[\hat{\theta}_l,\hat{\theta}_u]$ 表示一个连续型模糊总体参数 $F\theta=[l,u]$ 的估计量，若 $E(F\hat{\theta})=F\theta$，则称 $F\hat{\theta}$ 为连续型模糊总体参数 $F\theta$ 的模糊无偏估计量。

【性质 5.1】 设 $F\mu=\frac{m_1}{L_1}+\frac{m_2}{L_2}+\cdots+\frac{m_k}{L_k}$ 表示一个离散型模糊总体均数，则 $F\bar{X}$ 为 $F\mu$ 的模糊无偏估计量。

证明 若 $\left\{X_i=\frac{m_{i1}}{L_1}+\frac{m_{i2}}{L_2}+\cdots+\frac{m_{ik}}{L_k},i=1,2,\cdots,n\right\}$ 为从此模糊总体均数 $F\mu$ 中所抽取的一组随机样本，则

$$E(X_i)=\frac{E(m_{i1})}{L_1}+\frac{E(m_{i2})}{L_2}+\cdots+\frac{E(m_{ik})}{L_k}=\frac{m_1}{L_1}+\frac{m_2}{L_2}+\cdots+\frac{m_k}{L_k}$$

又
$$F\bar{X}=\frac{\frac{1}{n}\sum_{i=1}^{n}m_{i1}}{L_1}+\frac{\frac{1}{n}\sum_{i=1}^{n}m_{i2}}{L_2}+\cdots+\frac{\frac{1}{n}\sum_{i=1}^{n}m_{ik}}{L_k}$$
故
$$E(F\bar{X})=\frac{E\left(\frac{1}{n}\sum_{i=1}^{n}m_{i1}\right)}{L_1}+\frac{E\left(\frac{1}{n}\sum_{i=1}^{n}m_{i2}\right)}{L_2}+\cdots+\frac{E\left(\frac{1}{n}\sum_{i=1}^{n}m_{ik}\right)}{L_k}=$$
$$\frac{\frac{1}{n}\sum_{i=1}^{n}E(m_{i1})}{L_1}+\frac{\frac{1}{n}\sum_{i=1}^{n}E(m_{i2})}{L_2}+\cdots+\frac{\frac{1}{n}\sum_{i=1}^{n}E(m_{ik})}{L_k}=$$
$$\frac{m_1}{L_1}+\frac{m_2}{L_2}+\cdots+\frac{m_k}{L_k}=F\mu$$
$F\bar{X}$ 为 $F\mu$ 的模糊无偏估计量。

【性质5.2】　设一个连续型模糊总体平均数为 $F\mu=[l,u]$，则 $F\bar{X}$ 为 $F\mu$ 的模糊无偏估计量。

证明　若 $\{X_i=[X_{il},X_{iu}],i=1,2,\cdots,n\}$ 为此模糊总体均数 $F\mu$ 中的一组模糊随机变量，则 $E(X_i)=[E(X_{il}),E(X_{iu})]=[l,u]$。又由于
$$F\bar{X}=\left[\frac{1}{n}\sum_{i=1}^{n}X_{il},\frac{1}{n}\sum_{i=1}^{n}X_{iu}\right]$$
故
$$E(F\bar{X})=\left[E\left(\frac{1}{n}\sum_{i=1}^{n}X_{il}\right),E\left(\frac{1}{n}\sum_{i=1}^{n}X_{iu}\right)\right]=\left[\frac{1}{n}\sum_{i=1}^{n}E(X_{il}),\frac{1}{n}\sum_{i=1}^{n}E(X_{iu})\right]=$$
$$\left[\frac{1}{n}\sum_{i=1}^{n}l\ \frac{1}{n}\sum_{i=1}^{n}u\right]=[l,u]$$
证得，$F\bar{X}$ 是 $F\mu$ 的模糊无偏估计量。

【定义 5.14】　离散型相对有效(Efficient)估计量

设 $F\hat{\theta}=\frac{\hat{m}_1}{L_1}+\frac{\hat{m}_2}{L_2}+\cdots+\frac{\hat{m}_k}{L_k}$，$F\dot{\hat{\theta}}=\frac{\dot{\hat{m}}_1}{L_1}+\frac{\dot{\hat{m}}_2}{L_2}+\cdots+\frac{\dot{\hat{m}}_k}{L_k}$ 表示两个离散型模糊总体参数 $F\theta=\frac{m_1}{L_1}+\frac{m_2}{L_2}+\cdots+\frac{m_k}{L_k}$ 的估计量。若
$$d_1(F\hat{\theta},F\theta)\leqslant d_1(F\dot{\hat{\theta}},F\theta)$$
其中 $d_1(A,B)=\frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}(\mu_A(x_i)-\mu_B(x_i))^2}$（$n$ 为论域元素个数），则称 $F\hat{\theta}$ 比 $F\dot{\hat{\theta}}$ 是更有效的估计量。

第 6 章　模糊数据假设检验

古典的统计检验必须陈列明确的假设。比如，当我们想要检验两总体平均数是否有差异时，零假设是“两个平均数相等”。然而，有时我们想要知道的只是两平均值是否非常逼近，此时传统的检验方法并不适用于这种包含不确定性的假设检验。因此，本章提出基于模糊数据样本的统计检验方法，针对模糊均数相等、模糊属于以及卡方同质性检验作出详细探讨。

6.1　距　　离

在社会科学的分析和研究里，特别是当我们想要利用统计参数分析强调统计检验应用于决策时，在传统的统计分析中，均数检验仅是限于单一数值的统计量，无法表现出人类思维及主观意识的多变性，因此，为了要将人类思维作更合理且准确的描述，发展模糊数据均数检验是迫切需要的。

卡塞尔斯等人(Casals et. al, 1986, 1989) 曾以模糊事件表征的模型为基础来讨论统计假设检验；萨德和施瓦茨兰德(Saade and Schwarzlander, 1990) 针对混合数据(Hybrid Data) 提出模糊假设检验；宫本(Miyamoto, 1990, 第 240 页) 也曾论及模糊假设检验。但是，他们对模糊假设检验方法的构思，不但基于数据的随机不确定性，也包含了数据中的非随机不确定性。然而，“两总体平均数几乎相等” 的假设检验，只讨论包含于假设中的非随机不确定性，这样的不确定性不但与数据本身无关，且无法假设于数据上。因此，卡塞尔斯等人(1986, 1989)，萨德和施瓦茨兰德(1990)，甚至贝叶斯近似所提供的框架尚无法处理上述的假设检验。本章将对内曼和皮尔逊(Neyman and Pearson) 设立的古典假设检验与决策模型加以扩展，并提出模糊假设的模糊统计检验，而用此方法检验所得的推断结论也是模糊数据的。

模糊总体均数统计检验是一个崭新的研究专题。首先，我们考虑对离散型与连续型模糊数进行反模糊化。然后，根据检验法则，在模糊检验水平 α 下，进行单侧或双侧的模糊总体均数检验。这里的模糊检验水平 α，有别于传统的检验水平 α，依赖于模糊总体的特性，以更稳健性的数值考虑拒绝区域 F_α。

在阐述模糊区间集合的距离之前，首先回顾普通距离函数的定义。

在数学、统计学等学科的理论及应用研究中，有关距离的概念起着十分重要的作用。引入距离可用于反映和刻画两个点或者两个集合之间的密切程度，也就是它们之间的远近程度的测量。

普通距离函数

【定义 6.1】　设 U 表示一个非空的集合，d 为定义在 U 上的一个二元函数，也就是

$$d: U \times U \to \{0\} \cup \mathbf{R}^+ = [0, \infty) \tag{6.1}$$
$$(u, v) \to d(u, v)$$

若 d 满足下述三个公理：

(1) 非负性

$$d(u,v) \geqslant 0, d(u,v) = 0 \leftrightarrow u = v, \forall u,v \in U \tag{6.2}$$

(2) 对称性

$$d(u,v) = d(v,u), \forall u,v \in U \tag{6.3}$$

(3) 三角不等式

$$d(u,w) \leqslant d(u,v) + d(v,w), \forall u,v,w \in U \tag{6.4}$$

则称 d 为 U 上的一个距离函数或者度量函数，称 $d(u,v)$ 为 u 与 v 两点之间的距离，而把由 U 与 d 构成的序对 (U,d) 称为距离空间或度量空间。

由定义 6.1 知，一个非负实数 $d(u,v)$ 成为 U 中两点 u 与 v 的距离函数当且仅当满足非负性、对称性、三角不等式三个公理。

【例 6.1】　设 $U = \{0\} \cup \mathbf{R}^+ = [0, \infty)$，定义 $d: R \times R \to \mathbf{R}^+$ 为

$$d(u,v) = |u - v|, \forall u,v \in U \tag{6.5}$$

容易验证，$d(u,v)$ 满足距离定义的三个公理，因此 d 是 R 上的一个距离函数，从而 $(\mathbf{R}^+, d)$ 构成一个距离空间，这是通常的一维实数空间。

【例 6.2】　设 $U = \mathbf{R}^n$ 表示全体维向量所构成的集合，定义 $d: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^+$ 为

$$d_2(u,v) = \left(\sum_{i=1}^{n} |u_i - v_i|^2\right)^{1/2}, \forall u,v \in \mathbf{R}^n \tag{6.6}$$

其中 $u = (u_1, u_2, \cdots, u_n)$，$v = (v_1, v_2, \cdots, v_n)$，容易验证，$d_2(u,v)$ 满足距离定义的三个公理，因此 d 是 $\mathbf{R}^n$ 上的一个距离函数，称为欧几里得(Euclid)距离，从而 $(\mathbf{R}^n, d_2)$ 称为欧几里得空间或欧氏空间。

实际上，一个空间可以定义出各种不同的距离。比如，上面例 6.2 中，还可以把距离定义成为

$$d_p(u,v) = (\sum_{i=1}^{n} | u_i - v_i |^p)^{1/p}, 1 \leqslant p < \infty, \forall u, v \in \mathbf{R}^n \quad (6.7)$$

而当令 $p = 2$ 时，就变成欧几里得距离。

一旦具有点到点的距离定义，就可以定义出点到集合、集合到集合之间的距离。下面，分别给出这两个定义。

【定义 6.2】 设 (U,d) 表示度量空间，U 为非空集合，把

$$d(x,A) = \inf_{v \in A} d(x,v) \quad (6.8)$$

称为度量空间 (U,d) 中点 x 到集合 A 的距离。

【定义 6.3】 设 (U,d) 表示度量空间，U 为非空集合，把

$$d(A,B) = \inf_{u \in A, v \in B} d(u,v) \quad (6.9)$$

称为度量空间 (U,d) 中集合 A 到集合 B 的距离。而把

$$d_H(A,B) = \sup_{u \in A, v \in B} \{d(u,B), d(v,A)\} \quad (6.10)$$

称为度量空间 (U,d) 中集合 A 到集合 B 的豪斯道夫(Hausdorff)距离。

模糊数之间距离

由于模糊数是对通常数的一种推广形式，所以也可以类似于度量空间中定义距离那样来定义模糊数之间的距离。

现在给出模糊数之间距离的定义。

【定义 6.4】 模糊区间集合的距离

设 A 表示具有隶属函数 $\mu_A(x) = f(x)$，其支集(Support)是模糊区间集合 $[a_l, a_u]$；B 表示具有隶属函数 $\mu_B(y) = g(y)$，其支集是模糊区间集合 $[b_l, b_u]$。下面，我们定义三种距离

$$d_1(A,B) = \inf\{| a - b | \mid a \in A, b \in B\} \quad (6.11)$$

$$d_2(A,B) = \sup\{| a - b | \mid a \in A, b \in B\} \quad (6.12)$$

$$d_3(A,B) = \inf\{| a_l - b_l |, | a_u - b_u |\} \quad (6.13)$$

【定义 6.5】 模糊等于

设 A 表示具有隶属函数 $\mu_A(x) = f(x)$，其支集是模糊区间集合 $[a_l, a_u]$；B 表示具有隶属函数 $\mu_B(y) = g(y)$，其支集是模糊区间集合 $[b_l, b_u]$。若 $[a_l, a_u] = [b_l, b_u]$，则称 A 模糊等于 B，写成 $A =_F B$。类似地，可以给出无上界区间与无下界区间的模糊等于。

【定义 6.6】 模糊属于

设 A 表示具有隶属函数 $\mu_A(x) = f(x)$，其支集是模糊区间集合 $[a_l, a_u]$；B 表示具有隶属函数 $\mu_B(y) = g(y)$，其支集是模糊区间集合 $[b_l, b_u]$。若

$[a_l, a_u] \subset [b_l, b_u]$，则称 A 模糊属于 B，写成 $A \in_F B$。类似地，可以给出无上界区间与无下界区间的模糊属于。

6.2　模糊数据总体均数检验

设 $F\mu$ 表示模糊数据样本总体均数，我们想要检验在模糊检验水平 α 下，是否接受 $H_0: F\mu = F\mu_0$ 的假设，其中 $F\mu_0$ 表示模糊总体均数。

离散型模糊总体均数检验方法

1. 样本：设 Ω 表示一个论域，令 $\{L_j, j = 1, \cdots, k\}$ 为论域 Ω 上的 k 个语言变量，$\{x_i, i = 1, \cdots, n\}$ 为一组模糊随机样本，且对每个样本 x_i 相对于语言变量 L_j 有一标准化的隶属度 m_{ij}。令 $L_{max} - L_{min}$ 表示相对于语言变量的全距（以 5 等级为例，即为 $5 - 1 = 4$）。

2. 统计假设 $H_0: F\mu = F\mu_0$ 对 $H_1: F\mu \neq F\mu_0$。

3. 统计量：求此组模糊样本 $\{x_i, i = 1, \cdots, n\}$ 的模糊样本均数 $F\bar{X}$。计算样本均数与总体均数的反模糊化值 $\bar{X}_F$ 与 μ_0。

4. 决策准则：在模糊检验水平 α 下，若 $|\bar{X}_F - \mu_0| > \alpha(L_{max} - L_{min})$，则拒绝 H_0。

注意到，对于左侧检验 $H_0: \mu \leqslant \mu_0$ 对 $H_1: \mu > \mu_0$ 在模糊检验水平 α 下，决策法则为：若 $\bar{X}_F - \mu_0 > \alpha(L_{max} - L_{min})$，则拒绝 H_0。类似地，可以给出右侧检验的情况。

关于连续型模糊样本，为了使符号一致，我们直接说 $F\mu$ 等于某区间，而不一定把它反模糊化。

连续型模糊总体均数检验方法

1. 样本：设 Ω 表示一个具有模糊均数 $[a, b]$ 的论域，令 $\{x_i = [x_{li}, x_{ui}], i = 1, \cdots, n\}$ 为一组模糊区间随机样本。

2. 统计假设 $H_0: F\mu =_F [a, b]$ 对 $H_1: F\mu \neq_F [a, b]$。

3. 统计量：计算 $F\bar{X} = [\bar{x}_l, \bar{x}_u]$。

4. 决策准则：在模糊检验水平 α 下，计算 $k = \alpha(b - a)$，若 $|\bar{x}_l - a| > k$ 或 $|\bar{x}_u - b| > k$，则拒绝 H_0。

进行区间相等的检验时，有时会遇到统计量落入事先假定的区间内的情况，但是可能由于区间相对过小，得到拒绝等于的结论。因此，我们必须考虑模

糊属于检验,以符合实际需要。

有界样本的模糊属于检验

1. 统计假设 $H_0: F\mu \in_F F\mu_0$ 对 $H_1: F\mu \notin_F F\mu_0$。

2. 统计量:随机抽取一组模糊样本$\{x_i, i = 1,\cdots,n\}$,计算 $F\bar{X} = [\bar{x}_l, \bar{x}_u]$。

3. 决策准则:在模糊检验水平 α 下,计算 $k = \alpha(b - a)$,若 $\bar{x}_l < a - k$ 或 $\bar{x}_u > b + k$ 时,拒绝 H_0。

无下界样本(Sample with No Lower Bound) 的模糊属于检验

1. 统计假设 $H_0: F\mu \in_F (-\infty, b)$ 对 $H_1: F\mu \notin_F (-\infty, b)$。

2. 统计量:随机抽样一组模糊样本$\{x_i, i = 1,\cdots,n\}$,计算 $F\bar{X} = (-\infty, \bar{x}_u)$。

3. 决策准则:在模糊检验水平 α 下,令 $k = \alpha b$,若 $x_u > b + k$ 时,拒绝 H_0。

无上界样本(Sample with No Upper Bound) 的模糊属于检验

1. 统计假设 $H_0: F\mu \in_F [a, \infty]$ 对 $H_1: F\mu \notin_F [a, \infty]$。

2. 随机抽样一组模糊样本$\{x_i, i = 1,\cdots,n\}$,计算 $F\bar{X} = [\bar{x}_l, \infty]$。

3. 决策准则:在模糊检验水平 α 下,令 $k = \alpha a$,当 $x_l < a - k$ 时拒绝 H_0。

【例 6.3】 哈尔滨市某农场主人想引进一种新品种的鸡,作为炸鸡用途。只要大部分人试吃后的平均评价在普通以上,他就决定引进大量繁殖。于是,他随机找出 5 位顾客试吃,然后依据语言变量给予隶属度,数据整理如表 6.1 所示。

表 6.1 农场主人想引进品种鸡的满意隶属函数

试吃者	1 = 很不满意	2 = 不满意	3 = 普通	4 = 满意	5 = 很满意
A	0	0	0	0.3	0.7
B	0	0	0	0	1
C	0	0.4	0.6	0	0
D	0	0	0	0.8	0.2
E	0.1	0.9	0	0	0

我们将此问题化成假设检验 $H_0: \mu_f \leqslant 3$ 对 $H_1: \mu_f > 3$。经计算得,$\bar{X}_f = 3.68$。在模糊检验水平 $\alpha = 0.1$ 下,$k = 0.1 \cdot (5 - 1)$。因为 $\bar{X}_f - \mu_f = 0.68 > 0.4$,故拒绝 H_0,因此,他决定引进该新品种的鸡。

【例 6.4】　某人想开服装店分店，准备只卖 12 种款式衣服。因为是区域性的，故他只找住在想要开店的地方附近的人员进行市场调查。他分别就 12 种款式衣服，随机各找出 10 人，给予 1 至 5 等级的满意度隶属度评分。他决定整体平均分数若大于 2.5 则开店，否则放弃开店。经过统计得 $\bar{A}_f = 3$，$\bar{B}_f = 2$，$\bar{C}_f = 3.1$，$\bar{D}_f = 2.5$，$\bar{E}_f = 2.6$，$\bar{F}_f = 4$，$\bar{G}_f = 1.2$，$\bar{H}_f = 2$，$\bar{I}_f = 1.8$，$\bar{J}_f = 2.9$，求出

$$\bar{X}_f = \frac{1}{10}(3 + 2 + 3.1 + 2.5 + 2.6 + 4 + 1.2 + 2 + 1.8 + 2.9) = 2.51$$

在模糊检验水平 $\alpha = 0.1$ 下，$k = 0.1 \cdot 4 = 0.4$。因为 $2.51 - 2.5 = 0.01$，而 $0.01 < 0.4$，所以他决定不开店。

【例 6.5】　南京市某大公司人力资源部提出，现今 20 岁至 26 岁的年轻人要求平均年薪待遇为 2 万至 4 万元。统计主管想要检验此报告是否属实，于是，随机找出 6 位 20 岁至 26 岁的年轻人调查，得到他们所要求的待遇分别为

$$[3,4],[1.8,2],[2,3],[4,6],[2,2.5],[2.5,3]$$

统计假设 $H_0: F\mu =_F [2,4]$ 对 $H_1: F\mu \neq_F [2,4]$。

根据模糊样本均数定义可得

$$F\bar{X} = [\bar{x}_l, \bar{x}_u] = \left[\frac{3 + 1.8 + 2 + 4 + 2 + 2.5}{6}, \frac{4 + 2 + 3 + 6 + 2.5 + 3}{6}\right] = [2.55, 3.42]$$

在模糊检验水平 $\alpha = 0.1$ 下，计算 $k = 0.1 \cdot (4 - 2) = 0.2$。因为 $2.55 > 2 \pm 0.2$，$3.42 < 4 \pm 0.2$，所以拒绝人力资源部平均年薪待遇为 2 万至 4 万元的说法。

但是，$[2.46, 3.42]$ 确实落于 $[2,4]$ 区间，$F\mu \in_F [2,4]$，也就是 20 岁至 26 岁的年轻人要求平均年薪待遇区间属于 2 万至 4 万元区间。

【例 6.6】　厦门市某家公司想让员工有舒适的工作环境以增加工作效率，只要员工觉得有点热，空调就启动。主管认为 29 ℃ 以上就觉得热，他不知道是否大家都这么认为。于是，他随机找出 5 位员工来调查，得到 5 个数据：28 ℃ 以上，26 ℃ 以上，30 ℃ 以上，27 ℃ 以上，26 ℃ 以上；亦即 $[28, \infty]$，$[26, \infty]$，$[30, \infty]$，$[27, \infty]$，$[26, \infty]$。

统计假设 $H_0: \mu =_F [28, \infty]$ 对 $H_1: \mu \neq_F [28, \infty]$。

计算得到，$F\bar{X} = [\bar{x}_l, \bar{x}_u] = [27.4, \infty]$。在模糊检验水平 $\alpha = 0.05$ 下，计算 $k = 0.05 \cdot 28 = 1.4$。因为 $27.4 - 29 = -1.6 < -1.4$，所以拒绝接受该主管的想法。若可改成 28 ℃ 就开启空调，则 $27.4 - 28 = -0.6 > -1.4$，比较容易被大家接受。

6.3 模糊分类数据的卡方同质性检验

卡方同质性检验用来决定两个或两个以上的总体中，各分类的比例是否同质的统计检验方法。我们使用的卡方同质性检验统计量为

$$\chi^2 = \sum_{i=A,B}\sum_{j=1}^{c}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} \tag{6.14}$$

其中，n_{ij} 表示第 i 组样本，第 j 类的观察次数，e_{ij} 表示在 H_0 下第 i 组样本，第 j 类的期望次数。当 n 足够大时，卡方同质性检验统计量 χ^2 会渐近服从自由度为 $(c-1)$ 的 χ^2 分布。因此，在显著水平 α 下，可查得 χ^2 分布的临界值 $\chi^2_\alpha(c-1)$，若检验统计量 χ^2 值大于 $\chi^2_\alpha(c-1)$，则拒绝 H_0。

若从选项隶属度来考虑，分类数据的单位是可再分割的。例如，当我们被问到对政府某项施政满意度时，可能是 0.6 满意 0.4 非常满意的模糊样本。此时，传统 χ^2 检验便无法处理此类分类数据问题。因此，为了解决此类问题，我们提出模糊分类数据的卡方 χ^2 检验过程如下。

离散型模糊总体均数同质性检验

1. 样本：设 Ω 表示一个论域，令 $\{L_j, j=1,\cdots,k\}$ 为论域 Ω 上的 k 个语言变量，$\{a_1,a_2,\cdots,a_m\}$ 与 $\{b_1,b_2,\cdots,b_n\}$ 来自两不同模糊总体 A，B 两组模糊随机样本。且每个随机样本相对于语言变量 L_j 均有一标准化的隶属度 mA_{ij}，mB_{ij}。$MA_j = \sum_{i=1}^{m} mA_{ij}$，$MB_j = \sum_{i=1}^{n} mB_{ij}$，为样本对语言变量 L_j 的隶属度总和。

2. 事先假设 H_0: $F\mu_A = {}_F F\mu_B$，A，B 两总体有相同的分布比率。

$$F\mu_A = \frac{\frac{MA_1}{m}}{L_1} + \frac{\frac{MA_2}{m}}{L_2} + \cdots + \frac{\frac{MA_k}{m}}{L_k},\ F\mu_B = \frac{\frac{MB_1}{n}}{L_1} + \frac{\frac{MB_2}{n}}{L_2} + \cdots + \frac{\frac{MB_k}{n}}{L_k}$$

3. 统计量：$\chi^2 = \sum_{i=A,B}\sum_{j=1}^{k}\frac{([Mi_j]-e_{ij})^2}{e_{ij}}$（$e_{ij}$ 表示期望次数，为了符合软件计算 χ^2 检验要求，我们表格的各单元隶属度总和用 4 舍 5 入以取得整数值。对于样本量大于 25 个的模糊样本来说，其结果对决策影响并不大）。

4. 决策法则：在显著水平 α 下，若 $\chi^2 > \chi^2_\alpha(k-1)$，则拒绝 H_0。

区间型模糊总体均数同质性检验

1. 样本：设 Ω 表示一个论域，令 $\{L_j, j=1,\cdots,k\}$ 为论域 Ω 上的 k 个语言变

量，$\{a_1, a_2, \cdots, a_m\}$ 与 $\{b_1, b_2, \cdots, b_n\}$ 来自两个不同模糊总体 A，B 两组模糊区间随机样本。且每个随机样本相对于语言变量 L_j 均有一标准化的隶属度 mA_{ij}，mB_{ij}。$MA_j = \sum_{i=1}^{m} mA_{ij}$，$MB_j = \sum_{i=1}^{n} mB_{ij}$，为样本对语言变量 L_j 的隶属度总和。

2. 事先假设 $H_0: F\mu_A = {}_F F\mu_B$，$A$，$B$ 两总体有相同的分布比率。

3. 统计量：$\chi^2 = \sum_{i=A,B} \sum_{j=1}^{k} \frac{([Mi_j] - e_{ij})^2}{e_{ij}}$（$e_{ij}$ 表示期望次数，为了符合软件计算 χ^2 检验要求，我们表格的各单元隶属度总和用 4 舍 5 入以取得整数值）。

4. 决策法则：在显著水平 α 下，若 $\chi^2 > \chi^2_\alpha(k-1)$，则拒绝 H_0。

【例 6.7】　某政党竞选总部在首尔市选举期间分析选情，想了解选民的性别对政党、政营的支持度比率是否相同，于是委托选举研究中心探讨选民的性别对政党的支持度比率是否有差异，进行两种问卷调查：1. 传统勾选一项；2. 模糊隶属度选项。其结果如表 6.2 所示。

表 6.2　选民的性别对政党、政营的支持度

	政党的支持度			χ^2 同质性检验	政党的模糊支持度			χ^2 同质性检验
类别	A 党	B 党	其他	$\chi^2 = 13.43 >$	A 党	B 党	其他	$\chi^2 = 2.41 <$
男性	220	280	100	$5.99 =$	256.3	229.4	114.2	$5.99 =$
女性	180	140	80	$\chi^2_{0.05}(2)$	151.7	161.2	87.1	$\chi^2_{0.05}(2)$

统计假设 H_0：选民的性别对政党的支持度比率相同。H_1：选民的性别对政党的支持度比率不相同。其中，查卡方分布附表知，$\chi^2_{0.05}(2) = 5.99$。

在显著水平 $\alpha = 0.05$ 下，可以看出选民性别对政党的支持度比率结果的差异。若应用传统回答法，则 χ^2 同质性检验结果为拒绝 H_0。而若用模糊隶属度回答，则 χ^2 同质性检验结果为接受 H_0。由此可看出，在用累加模糊隶属度回答过程中，把隐性隶属度考虑进去，会造成总隶属度值与仅投一票方法的差异。

【例 6.8】　哈尔滨市某家乐福大卖场想比较 X，Y 两个小区哪一区每月生活费支出较高，以便有针对性地制定不同的生鲜市场促销策略。他们在两小区各随机找 50 户调查。填表过程允许客户以隶属函数代替传统只能勾选一项。例如，客户可填，3 万元 ~ 5 万元为 0.7，5 万元 ~ 8 万元为 0.3 等。然后，把每户加总后，得到表 6.3。

表 6.3 X, Y 两小区每户每月生活费支出

	3万元以下	3万元 ~ 5万元	5万元 ~ 8万元	8万元 ~ 12万元	12万元以上
X 小区	2.8	10.3	19.7	14.2	5.0
Y 小区	7.1	21.6	20.9	6.8	2.6

统计假设 H_0: X, Y 两小区每月支出比率一致；H_1: Y 小区每月生活费支出较 X 小区高。

经计算得到，$\chi^2 = 8.43 > \chi^2_{0.1}(4) = 7.78$，其中查卡方分布附表知，$\chi^2_{0.1}(4) = 7.7794 \approx 7.78$。因此，在显著水平 $\alpha = 0.1$ 下，我们拒绝 H_0: X, Y 两小区每月支出比率一致。因而，可以说，Y 小区每月生活费支出较 X 小区高。

第 7 章　模糊聚类分析

传统的统计分类方法是根据一个适当的统计量，把总体分成若干个类（Clusters，又称为群、群落、群聚、组等）。但是，有时由于考虑的变量因子太多，还有基于二分法的逻辑思维，分类效果往往不理想。另外，对论域各因子的权数比重，在传统的分类方法里，亦未提及。因此，如何应用模糊分类方法，如何决定权数比重，以期得到更完善的分类效果，就是本章研究的重点内容。

本章首先介绍模糊分类算法，接着考虑依据隶属度的观念定义模糊权重及模糊相对权重，借助于因子特性的加重计算，改进传统聚类方法。同时，特别强调对于评价因素为多变量时，加入模糊权重的分布与给定，使分类结果更为合理化，提高分类效率与结果。

当统计数据为分类数据形式时，通常以聚类分析方式处理。聚类分析考虑总体中具有相似特性的类，将其归并在一起。在多变量分析与模式识别（Pattern Recognition）中起着十分重要的作用。聚类分析的应用，在许多领域中，比如金融、经济、生物科技、医学诊断、地质天文、教育心理、商业、管理等，均受到相当重视。

近年来，模糊聚类分析被广泛地应用在各个学科领域。例如，利用模糊聚类分析对学生在解答物理问题的想象力上作多因子的评价。在统计推断上采用模糊分割的理念及可能性理论。还有，对水质作分析，以观测生物上的大肠杆菌，但是因其未加入权重因素，以致调整的次数过大，在大量重复计算上浪费较多时间。一些学者提出，模糊权重相加法则及其注释，但对于权重的如何决定仍未说明，或者仍然延用过去主观经验的定义方式。

7.1　模糊聚类法

给定一组数据 X，聚类分析的目的是找出数个聚类中心（Clustering Centroids，又称类中心），将 X 分成数个适合的类。在传统聚类分析中，这些类中心被要求形成 X 的分割，并且使得在相同分割区块内的数据相关性较强，不同分割区块的数据之间相关性较弱。

聚类中心依据排序情况可分为三种：(1) 有序聚类中心（Ordered Cluster

Centroids)：聚类中心的散布呈递增或递减之排序状态；(2) 部分排序聚类中心(Partial Ordered Cluster Centroids)：聚类中心的散布只基于某些因素而呈排序情况；(3) 加权排序聚类中心(Weighted Ordered Cluster Centroids)：聚类中心加入权重的因素，使其散布情况依据因素与权重排序的状态而出现。

传统上，我们经常使用 k 平均法(k - means) 进行数据分类，其分类步骤为：(1) 设聚类个数为 c，并计算各类的聚类中心。(2) 计算每一样本到各中心之距离，并将各样本分配到与其最近的类。(3) 重新计算新聚类中心 v_i。(4) 重复步骤(2) 与(3)，直到各类没有重新分配样本的情况为止。

在模糊聚类分析中有两个基本方法，其中之一是依据模糊 c 分类，被称为模糊 c 平均分类法。另一个方法是根据模糊等价关系计算，称为基于模糊等价关系聚类分类法(Clustering Method Based upon Fuzzy Equivalence Relation)。下面将分别介绍这两种方法，并用事例加以阐明。

模糊分类法中的模糊 c 平均法，除了必须事先指定聚类数 c，另外选定一个实数 m 与一个代表停止准则的微小正数 ϵ。与传统分类法最大的不同在于，隶属函数与特征函数的差别。显然，隶属函数的值域是介于 0 到 1 之间的所有实数。模糊聚类的含义为：给定一组数据 $X = \{x_1, x_2, \cdots, x_n\}$，把其分类成一组模糊子集 $P = \{P_1, P_2, \cdots, P_c\}$，且满足下列条件

$$\sum_{i=1}^{c} P_i(x_j) = 1, \text{对所有 } j \in \mathbf{N}, \text{且 } 0 < P_i(x_j) < 1$$

【例 7.1】 设一组数据集 $X = \{x_1, x_2, x_3, x_4\}$，若 $P = \{P_1, P_2\}$ 为 X 的一个分割，其隶属情况，如下表所示：

X 的元素	x_1	x_2	x_3	x_4
属于 P_1 的隶属度	0.2	0.9	0.6	0
属于 P_2 的隶属度	0.8	0.1	0.4	1

$$P_1 = 0.2I_{x_1} + 0.9I_{x_2} + 0.6I_{x_3} + 0I_{x_4}, P_2 = 0.8I_{x_1} + 0.1I_{x_2} + 0.4I_{x_3} + 1I_{x_4}$$

这里的 $P = \{P_1, P_2\}$ 就是一个模糊 2 分类。

给定一组数据集合 $X = \{x_1, x_2, \cdots, x_n\}$，一般地说，$x_k$ 是一个向量。模糊聚类分析的目标是找出一组有 c 个聚类中心 $v_1, v_2, \cdots, v_c$ 的模糊分类 $P = \{P_1, P_2, \cdots, P_c\}$，这些聚类中心要尽可能清楚表示样本分布情况。我们需要一些规则来表达这个概念，也就是对结果能使同一个聚类中的元素有较强的关联，且不同的聚类中的元素关联性较弱的规则加以考虑。因此，定义一个模糊 c 分割矩阵$M_{fc} = \{P \mid \mu_{ij}\}$，$\mu_{ij}$ 表示样本j 属于i 聚类的隶属度，且满足式(7.1)，(7.2)以及式(7.3)

$$\mu_{ij} \in [0,1]; i = 1,\cdots,c; j = 1,\cdots,n \tag{7.1}$$

$$\sum_{i=1}^{c} \mu_{ij} = 1; j = 1,\cdots,n \tag{7.2}$$

$$0 < \sum_{j=1}^{n} \mu_{ij} < n \tag{7.3}$$

则模糊聚类的目标函数为

$$J_{fc}(P,v) = \sum_{j=1}^{n}\sum_{i=1}^{c}(\mu_{ij})^m \| x_j - v_i \|^2 \tag{7.4}$$

其中 $1 \leqslant m < \infty$ 表示模糊指数(通常,为了方便计算取 m 为正整数),此加权参数控制着分类过程的模糊性。$\| \cdot \|$ 表示此空间中的内积,$\| x_j - v_i \|^2$ 表示 x_j 与 v_i 的欧几里得距离。

在式(7.1),(7.2) 以及式(7.3) 的条件下,对式(7.4) 求最小值,可得 μ_{ij} 及 v_i 为

$$v_i = \frac{\sum_{j=1}^{n}(\mu_{ij})^m x_j}{\sum_{j=1}^{n}(\mu_{ij})^m}, i = 1,2,\cdots,c \tag{7.5}$$

$$\mu_{ij} = \frac{(1/\| x_j - v_i \|^2)^{1/(m-1)}}{\sum_{i=1}^{c}(1/\| x_j - v_i \|^2)^{1/(m-1)}}; i = 1,2,\cdots,c; j = 1,\cdots,n \tag{7.6}$$

注意到,由式(7.5) 计算出的 v_i,它可被视为模糊分类 P_i 的聚类中心,因为它是 P_i 中数据的加权平均,而权数是 x_j 在模糊集合 P_i 中隶属度的 m 次方。

目标函数用来测度模糊聚类中的聚类中心与元素距离加权后的和。所以,较小的 $J_m(P)$ 值对应较佳的模糊分类。模糊 c 平均分类法的目标是找出一组模糊分类 P 拥有最小的 $J_m(P)$;聚类问题转变成优化问题。

模糊 c 平均分类算法

步骤 1:选择 c 个起始类中心,起始模糊分类 $P^{(0)}$,m 以及误差值 ε。

步骤 2:由式(7.5) 计算 $P^{(t)}$ 的 c 个类中心 $v^{(t)} = \{v_1^{(t)}, v_2^{(t)}, \cdots, v_c^{(t)}\}$。

步骤 3:由 $v^{(t)} = \{v_1^{(t)}, v_2^{(t)}, \cdots, v_c^{(t)}\}$ 与式(7.6) 计算新的模糊分类

$$P^{(t+1)} = \{P_1^{(t+1)}, P_2^{(t+1)}, \cdots, P_c^{(t+1)}\}$$

步骤 4:比较 $P^{(t)}$ 与 $P^{(t+1)}$。若 $\| P^{(t)} - P^{(t+1)} \| < \varepsilon$,运算就停止。否则,令 $t = t + 1$,回到步骤 2。

在步骤 4 中,$\| P^{(t)} - P^{(t+1)} \|$ 表示 $P^{(t)}$ 与 $P^{(t+1)}$ 的距离。例如,可设

$$\| P^{(t)} - P^{(t+1)} \| = \max_{i=1,\cdots,c,j=1,\cdots,n} | \mu_{ij}^{(t+1)} - \mu_{ij}^{(t)} |$$

此算法中，m 的选取可根据问题的需要来考虑。当 m 愈接近 1，则分类结果愈趋近于传统分类法；当 m 趋近无限大，则 $J_{fk}^{(t)}$ 函数值将趋近于零，分类结果愈模糊。也就是，所有的聚类中心倾向于数据集 X 的中心。虽然文献上很少探讨应该如何选取 m 较为合适，但是根据经验上的最佳选择是介于范围 1.25 到 5 之间最为适当。目前，并无理论指出，如何选取最佳的 m 值，但对所有的 m，此算法会收敛。

【例 7.2】　金龙少棒队教练为了准备远东杯比赛的攻防战略考虑将 15 位小选手的体能状况 x_{k1}, x_{j1} 与打击率 x_{k2}, x_{j2} 作一个评估。下面为 15 位小选手的最近一次测验数据集合 X，如下表：

j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_{j1}	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.3	0.4	0.5	0.5	0.5	0.6	0.6	0.6
x_{j2}	0.0	0.2	0.4	0.1	0.2	0.3	0.2	0.2	0.2	0.1	0.2	0.3	0.0	0.2	0.4

假定我们要找 $c = 2$ 的模糊分类，并且设 $m = 1.25, \varepsilon = 0.01$，而 $\|\cdot\|$ 表示欧几里得距离。初始，给定模糊分类 $P^{(0)} = \{P_1, P_2\}$。

$$P_1 = 0.854I_{x_1} + 0.854I_{x_2} + \cdots + 0.854I_{x_{15}}$$

$$P_2 = 0.146I_{x_1} + 0.146I_{x_2} + \cdots + 0.146I_{x_{15}}$$

算法最后在 $t = 6$ 时停止，我们得到如下 $P^{(6)}$：

j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_{j1}	0.99	1	0.99	1	1	1	0.99	0.47	0.01	0	0	0	0.01	0	0.01
x_{j2}	0.01	0	0.01	0	0	0	0.01	0.53	0.99	1	1	1	0.99	1	0.99

与两个聚类中心 $v_1 = (0.09, 0.2)$ 与 $v_2 = (0.51, 0.2)$。

模糊等价关系聚类分类法

在使用模糊 c 平均分类法时，必须给定聚类中心的数目。当聚类问题并没有给定这一数目时，这是一个很不利的条件。在这种情况下，聚类中心的数目自然应由数据的结构反映出。依据模糊关系的分类法正是以这样的方法运作。

每个模糊关系对每一个分割都可以导出明确的分类。模糊聚类问题可视为适当地辨识数据模糊关系。虽然这通常无法直接作出，但我们可以应用合适的距离函数，就很容易确定定义良好的模糊关系。

如同前面所阐述的，给定一组数据集 $X = \{x_j \mid x_j \in R^p, j = 1, \cdots, n\}$。令在

X 上的定义良好模糊关系(Well-defined Fuzzy Relation)R，利用闵可夫斯基(Minkowski)分类上适当的距离函数，公式如下

$$R(x_i, x_k) = 1 - \delta\left(\sum_{l=1}^{p} |x_{il} - x_{kl}|^q\right)^{\frac{1}{q}} \tag{7.7}$$

其中 δ 表示一个常数，使 $R(x_i, x_k)$ 介于 0 与 1 之间。显然，δ 是 X 中最大距离的倒数。

一般地说，式(7.5)所定义的定义良好模糊关系，不一定是模糊等价关系。因此，我们通常必须决定 R 的传递闭包，这可用比较简单的算法完成。然而，如果 R 是定义良好模糊关系，我们可利用下述定理发展更有效率的算法。

【定理 7.1】　设 R 表示有限全集(Universal Set)$X(|X| = n)$ 上的定义良好模糊关系，则 R 的最大－最小的传递闭包集合是关系 $R^{(n-1)}$。

这个定理提供了计算传递闭包的方法，$R_T = R^{(n-1)}$，也就是，经由计算关系矩阵的数列

$$R^{(2)} = R \cdot R$$
$$R^{(4)} = R^{(2)} \cdot R^{(2)}$$
$$\vdots$$
$$R^{(2^k)} = R^{(2^{k-1})} \cdot R^{(2^{k-1})}$$

直到没有新的关系矩阵被算出，或 $2^k \geq n - 1$。这个算法在计算上比一般的算法更有效率，但只适用模糊反身关系而已。

【例 7.3】　给定一组数据集合 X 如下：

j	1	2	3	4	5
x_{j1}	0	1	2	3	4
x_{j2}	0	1	3	1	0

在此例中，为了看出式(7.5)中参数 q 的影响力，在分析数据时，取 $q = 1$，2。$q = 1$ 时，这时的距离函数为汉明距离。最大的距离是 5(x_1 与 x_3 的汉明距离)，取 $\delta = 1/5 = 0.2$。由式(7.5)，得到

$$R = \begin{bmatrix} 1 & 0.6 & 0 & 0.2 & 0.2 \\ 0.6 & 1 & 0.4 & 0.6 & 0.2 \\ 0 & 0.4 & 1 & 0.4 & 0 \\ 0.2 & 0.6 & 0.4 & 1 & 0.6 \\ 0.2 & 0.2 & 0 & 0.6 & 1 \end{bmatrix}$$

传递闭包为

$$R_T=\begin{bmatrix}1&0.6&0.4&0.6&0.6\\0.6&1&0.4&0.6&0.6\\0.4&0.4&1&0.4&0.4\\0.6&0.6&0.4&1&0.6\\0.6&0.6&0.4&0.6&1\end{bmatrix}$$

由此关系式,可找出不同 α 分割的分类如下

$\alpha\in[0,0.4]:\{\{x_1,x_2,x_3,x_4,x_5\}\}$

$\alpha\in(0.4,0.6]:\{\{x_1,x_2,x_4,x_5\},\{x_3\}\}$

$\alpha\in(0.6,1]:\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\}\}$

$q=2$时,这时的距离函数为欧几里得距离。我们的第一步是要决定:在 X 中,最大的欧几里得距离是4(x_1 与 x_5 两点的距离),故取 $\delta=1/4=0.25$。然后,由式(7.5)计算 R 的隶属度,例如,$R(x_1,x_3)=1-0.25(2^2+3^2)^{0.5}=0.1$,从而得到矩阵

$$R=\begin{bmatrix}1&0.65&0.1&0.21&0\\0.65&1&0.44&0.5&0.21\\0.1&0.44&1&0.44&0.1\\0.21&0.5&0.44&1&0.65\\0&0.21&0.1&0.65&1\end{bmatrix}$$

这关系矩阵不是最大 - 最小传递;它的传递闭包为

$$R_T=\begin{bmatrix}1&0.65&0.44&0.5&0.5\\0.65&1&0.44&0.5&0.5\\0.44&0.44&1&0.44&0.44\\0.5&0.5&0.44&1&0.65\\0.5&0.5&0.44&0.65&1\end{bmatrix}$$

当对应不同的 α 分割时,此关系矩阵导出4种不同的分类如下

$\alpha\in[0,0.44]:\{\{x_1,x_2,x_3,x_4,x_5\}\}$

$\alpha\in(0.44,0.5]:\{\{x_1,x_2,x_4,x_5\},\{x_3\}\}$

$\alpha\in(0.5,0.65]:\{\{x_1,x_2\},\{x_3\},\{x_4,x_5\}\}$

$\alpha\in(0.65,1]:\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\}\}$

下面实例是由田村等(Tamura, et al,1971)给出的,该例子具有额外条件的模糊等价关系分类法。

【例7.4】 假定我们有16幅来自3个家庭的肖像,以1到16(N_{16})来登记。若已知每个家庭包含4到7个成员。我们的目标是依据这些条件分类肖像。首先,要决定每对肖像的相似程度。收集一群人由某些方法作出个人的判断,这可

以是由主观来完成的。设相似程度已经得到了，那么它们可以被视为在 N_{16} 上的定义良好模糊关系 R 的隶属度，如表 7.1 所示。因为这矩阵是对称的，所以仅将下三角矩阵写出来。

在这个例子中，有一个有趣的特征；父母与孩子的相似性是被预期存在的，可是父亲与母亲之间并没有相似性。然而，经由 $\boldsymbol{R}$ 转换成它自己的传递闭包 $\boldsymbol{R}_T$，就可以建立父亲与母亲的相似性。应用定理 7.1 的算法，可得到 $\boldsymbol{R}_T = \boldsymbol{R}^{(8)}$，$\boldsymbol{R}$ 与 $\boldsymbol{R}_T$ 就是表 7.1(a) 与(b)。

表 7.1　(a) N_{16} 家庭的模糊关系矩阵 R

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1															
2	0	1														
3	0	0	1													
4	0	0	0.4	1												
5	0	0.8	0	0	1											
6	0.5	0	0.2	0.2	0	1										
7	0	0.8	0	0	0.4	0	1									
8	0.4	0.2	0.2	0.5	0	0.8	0	1								
9	0	0.4	0	0.8	0.4	0.2	0.4	0	1							
10	0	0	0.2	0.2	0	0	0.2	0	0.2	1						
11	0	0.5	0.2	0.2	0	0	0.8	0	0.4	0.2	1					
12	0	0	0.2	0.8	0	0	0	0	0.4	0.8	0	1				
13	0.8	0	0.2	0.4	0	0.4	0	0.4	0	0	0	0	1			
14	0	0.8	0	0.2	0.4	0	0.8	0	0.2	0.2	0.6	0	0	1		
15	0	0	0.4	0.8	0	0.2	0	0	0.2	0	0	0.2	0.2	0	1	
16	0.6	0	0	0.2	0.2	0.8	0	0.4	0	0	0	0	0.4	0.2	0	1

(b) N_{16} 家庭的模糊关系矩阵 R 的传递闭包 R_T

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1															
2	0.4	1														
3	0.4	0.4	1													
4	0.5	0.4	0.4	1												
5	0.4	0.8	0.4	0.4	1											
6	0.6	0.4	0.4	0.5	0.4	1										
7	0.4	0.8	0.4	0.4	0.8	0.4	1									
8	0.6	0.4	0.4	0.5	0.4	0.8	0.4	1								
9	0.5	0.4	0.4	0.8	0.4	0.5	0.4	0.5	1							
10	0.5	0.4	0.4	0.8	0.4	0.5	0.4	0.5	0.8	1						
11	0.4	0.8	0.4	0.4	0.8	0.4	0.8	0.4	0.4	0.4	1					
12	0.5	0.4	0.4	0.8	0.4	0.5	0.4	0.5	0.8	0.8	0.4	1				
13	0.8	0.4	0.4	0.5	0.4	0.6	0.4	0.6	0.5	0.5	0.4	0.5	1			
14	0.4	0.8	0.4	0.4	0.8	0.4	0.8	0.4	0.4	0.4	0.8	0.4	0.4	1		
15	0.5	0.4	0.4	0.8	0.4	0.5	0.4	0.5	0.8	0.8	0.4	0.8	0.5	0.4	1	
16	0.6	0.4	0.4	0.5	0.4	0.8	0.4	0.8	0.5	0.5	0.4	0.5	0.6	0.4	0.5	1

图 7.1 是 R_T 的树形图。我们发现，没有一组分类完全满足限制条件：三个家庭，每一家庭 4 到 7 人。满足最多限制条件的分类是对应 $\alpha = 0.8$ 时的 $R_T^{0.8} = \{\{1,3\},\{1,6,8,16\},\{2,3,5,7,11,14\},\{4,9,10,12,15\}\}$。

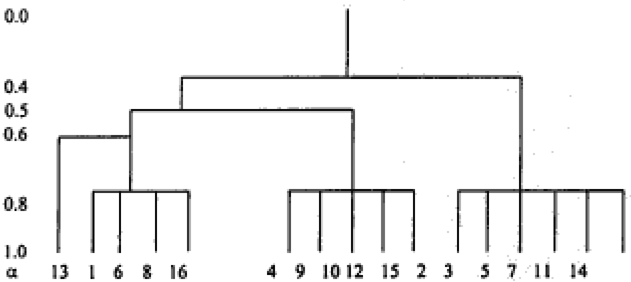


图 7.1 R_T 的树形图

在这些分类中，无法被满足的条件是肖像 3 无法被归类至任何一家庭。原因是肖像 3 是被收养或前次婚姻所留下的小孩肖像。

田村等人(Tamura, et al,1971)，在他们的报告中提到，他们作出把 60 个家庭分成 20 类、每类 3 个家庭的实验；结果，77% 可正确的分类，13% 分类错误，10% 无法分类。这令人印象深刻，因为在此都是十分模糊的信息与主观的判断。

7.2　模糊权重分析与判定程序

显然，若依据某一因素考虑样本时，它属于某一类；而依据另一因素考虑时，则可能属于另一类。可见，权重的决定及如何决定会影响到分类结果，因此，作多变量分类前我们必须要先对权重作一合理的估计。传统分类方法常把各因子以相同权重看待，但实际上，论域因子的权重应该不同，因为样本的特征因素未必同等重要。就像人们品茶时，特别重视茶的水色、香味、或口感因素，而决定选择何种茶叶。又比如，我们选购衣服时，先考虑衣服的价格、舒适度，还有颜色等等。可知不同的因子有着不同的权重，而对于各因子重视的程度，又因人而异，不易评估。若以相同权重看待各因子，或者以主观的个人意识作判断去定义因子权重，易产生不适当的分类结果。

如何决定各因子的权数?本章考虑用模糊统计法来进行对其估计。所谓模糊统计法就是将所有因子以模糊问卷的形式依据重要程度收集信息，并采用统计方法转换得到权重值。在社会调查中，人们常常被访问去陈述并预测他们未来的行为意图。

一般地说，通过对专家的访问或统计上的抽样调查，通常可以获得很好的预测效果。模糊统计法不同于以往的社会调查，在于传统的问卷总是使用某些固定的回答模式测度喜好程度，并以整数所衍生的名义尺度记录此评估测量与行为结果间的统计相关性。然而，在人类的思维与行为中，几乎都是反映事物模糊性的概念，所表现出来的自然语言也都是模糊语言。因此，对社会调查而言，模糊模式较适合于评估人类模糊思维的可能性(Possibility)及可行性(Feasibility)。

在说明模糊权重分析的运算之前，我们先对偏好序列、模糊权重以及模糊相对权重给出基本定义。

【定义 7.1】　偏好效用序列(Preference Utility Sequence)

假定偏好效用序列为 $r = \{r_1, r_2, \cdots, r_f\}$，则定义 $r_1 < r_2 < \cdots < r_f$ 为偏好

效用递增序列(Preference Utility Increasing Sequence);反之,$r_1 > r_2 > \cdots > r_f$ 为偏好效用递减序列(Preference Utility Decreasing Sequence)。

【定义 7.2】 模糊权重(Fuzzy Weight,FW)

假定论域集合 $S = \{S_1, S_2, \cdots, S_k\}$,偏好效用序列 $r = \{r_1, r_2, \cdots, r_f\}$,且 S_i 在 r_f 的隶属度为 μ_{S_f},则把论域因子的模糊权重 $FW = (FW_{S_1}, \cdots, FW_{S_k})$ 定义为

$$FW_{S_i} = \frac{\sum_{l=1}^{f} \mu_{S_i l}}{r_l} = \frac{\mu_{S_i 1}}{r_1} + \frac{\mu_{S_i 2}}{r_2} + \cdots + \frac{\mu_{S_i f}}{r_f}, i = 1, \cdots, c \tag{7.8}$$

下面,我们举一个简单例子阐明偏好序列与模糊权重中隶属函数所代表的意义。

【例7.5】 假定评估哈尔滨市区房屋价格的论域为 $S = \{S_1, S_2, S_3, S_4\}$ = {地点、格局大小、建材、外围设施},偏好效用序列为 $r = \{r_1, r_2, \cdots, r_5\}$ = {很不重要、不重要、普通、重要、很重要}。因为 $r_1 < r_2 < \cdots < r_5$,所以偏好效用序列 r 为偏好递增序列。

根据定义 7.2,倘若评估房屋价格时,衡量地点因素是“10% 的很不重要”和“37% 的不重要”,而衡量格局大小因素时是“20% 的重要” 和“80% 的很重要” 等等。以隶属函数表示,则为 $\mu_{S_1 1} = 0.1, \mu_{S_1 2} = 0.37, \mu_{S_2 4} = 0.2$ 以及 $\mu_{S_2 5} = 0.8$。

在模糊集合中,隶属度的范围从0到1。每一个语言变项,例如,形状代表一个可能性分布,而且关于分布的评定结果往往会因人而异。因而,将这些受访者的回答加以平均,以得到论域因子 S 在偏好序列 r 的隶属度 μ_S 较合理的分布。

模糊权重表示各个因子的自我权重分布,但是模糊权重分析的主要目的是求得因子所代表的权重值各是多少,亦即相对的权重。在这个基础上,进一步地定义模糊相对权重,以供我们在作模糊权重分析时使用。

【定义 7.3】 模糊相对权重(Fuzzy Relative Weight,FRW)

假定论域集合 $S = \{S_1, S_2, \cdots, S_k\}$,偏好效用序列 $r = \{r_1, r_2, \cdots, r_f\}$,且 S_i 在 r_j 的隶属函数为 μ_{S_j},则模糊相对权重 $FRW = (FRW_{S_1}, \cdots, FRW_{S_c})$ 为由模糊权重 FW 采用 m 等级评分标准法转换所得。若 $r_1 < r_2 < \cdots < r_f$,则

$$FRW_{S_i} = \frac{\sum_{l=1}^{f} l \cdot \mu_{S_i l}}{\sum_{i=1}^{k}\sum_{l=1}^{f} l \cdot \mu_{S_i l}}, i = 1, \cdots, k \tag{7.9}$$

反之,若 $r_1 > r_2 > \cdots > r_f$ 时,则

$$FRW_{S_i} = \frac{\sum_{l=1}^{f}(n-l+1)j \cdot \mu_{S_il}}{\sum_{i=1}^{k}\sum_{l=1}^{f}(n-l+1)\cdot \mu_{S_il}}, i=1,\cdots,k \tag{7.10}$$

为了清楚理解 FRW 从 FW 经由 f 等级评分标准法转换而来的过程，下面我们举例说明它们的关系。

【例 7.6】　假定人们在选购衣服时，考虑是否购买的因素有衣服的样式、质料、大小、颜色以及价格。而且，衡量的重要性各为：不重要、普通与重要。根据模糊权重的定义，得到 $k=7$ 与 $f=3$。也就是说，论域集合为 $S=\{S_1,S_2,S_3,S_4,S_5\}=\{$样式、质料、大小、颜色、价格$\}$，而偏好效用序列为 $r=\{r_1,r_2,r_3\}=\{$不重要、普通、重要$\}$。因为 $f=3$，故根据定义 7.3 利用三等级评分标准法去求出样式、质料、大小、颜色以及价格的模糊相对权重。

为了求论域因子的模糊相对权重，首先设计一个模糊评估表（Fuzzy Evaluation Table）的问卷形式如表 7.2 所示。

表 7.2　模糊评估表在偏好效用序列 U 的有序程度 μ_A

S \ r	r_1	r_2	$\cdots$	r_f
S_1	μ_{S_11}	μ_{S_12}	$\cdots$	μ_{S_1f}
S_2	μ_{S_21}	μ_{S_22}	$\cdots$	μ_{S_2f}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
S_k	μ_{S_k1}	μ_{S_k2}	$\cdots$	μ_{S_kf}

假定 $r_1 < r_2 < \cdots < r_f$，以此问卷访问专家对论域集合在各偏好序列下的有序程度。这些专家评分平均后的各因子的偏好度，即为模糊权重集。根据定义 7.2 知，论域的各因子的模糊权重分别为

$$FW_{S_1} = \sum_{l=1}^{f}\frac{\mu_{S_1l}}{r_l} = \frac{\mu_{S_11}}{r_1} + \frac{\mu_{S_12}}{r_2} + \cdots + \frac{\mu_{S_1f}}{r_f}$$

$$FW_{S_2} = \sum_{l=1}^{f}\frac{\mu_{S_2l}}{r_l} = \frac{\mu_{S_21}}{r_1} + \frac{\mu_{S_22}}{r_2} + \cdots + \frac{\mu_{S_2f}}{r_f}$$

$$\vdots$$

$$FW_{S_k} = \sum_{l=1}^{f}\frac{\mu_{S_kl}}{r_l} = \frac{\mu_{S_k1}}{r_1} + \frac{\mu_{S_k2}}{r_2} + \cdots + \frac{\mu_{S_kf}}{r_f}$$

再采用 m 等级评分标准法，分别计算论域中各因子的模糊相对权重 FRW。

所谓 m 等级评分标准法，就是将 m 个偏好序列 r 视为 f 个等级，对此 f 个偏好序列取数量化。也就是，给定 r_1 为一分，给定 r_2 为二分，如此继续到给定 r_f 为 f 分。根据所得的隶属度乘上其相对应的分数，分别求出它们相对的模糊权重分布，则为各因子的模糊相对权重。因此，以定义 7.3 的表示法，各因子的模糊相对权重 FRW 分别为

$$FRW_{S_1}=\frac{\sum_{l=1}^{f} l\cdot\mu_{S_1 l}}{\sum_{i=1}^{k}\sum_{l=1}^{f} l\cdot\mu_{S_i l}}$$

$$FRW_{S_2}=\frac{\sum_{l=1}^{f} l\cdot\mu_{S_2 l}}{\sum_{i=1}^{k}\sum_{l=1}^{f} l\cdot\mu_{S_i l}}$$

$$\vdots$$

$$FRW_{S_k}=\frac{\sum_{l=1}^{f} l\cdot\mu_{S_k l}}{\sum_{i=1}^{k}\sum_{l=1}^{f} l\cdot\mu_{S_i l}}$$

以上介绍的方法，是进行多变量分类时考虑各因子的模糊相对权数法。举一个关于茶叶方面的例子，以解释使用模糊权重分析法决定评估茶因子的权重集。

【例 7.7】 考虑以模糊评估表访问品茶专家在评定茶叶质量好坏时，衡量茶叶所采用的评价因素的重要程度。假定访问数位茶叶专家对模糊评估表评分，经过统计平均后的数据如表 7.3 所示。

表 7.3 平均后的偏好度

评估项目＼偏好序列	很不重要	不重要	普通	重要	很重要
形　状	0.10	0.35	0.50	0.55	0.65
色　泽	0.30	0.50	0.60	0.20	0.10
水　色	0.20	0.45	0.80	0.50	0.30
香　味	0	0.10	0.40	0.80	0.92

故根据定义 7.2，我们得到，这四种因子的模糊权重分别为

$FW_{形状}$ = 0.1/ 很不重要 + 0.35/ 不重要 + 0.5/ 普通 + 0.55/ 重要 + 0.65/ 很重要

$FW_{色泽}$ = 0.3/很不重要 + 0.5/不重要 + 0.6/普通 + 0.2/重要 + 0.1/很重要

$FW_{水色}$ = 0.2/很不重要 + 0.45/不重要 + 0.8/普通 + 0.5/重要 + 0.3/很重要

$FW_{香味}$ = 0.1/不重要 + 0.4/普通 + 0.8/重要 + 0.92/很重要

再利用五等级评分标准法，给定很重要为5分，重要为4分，普通为3分，不重要为2分，而很不重要则给定为1分。根据表7.3上的隶属度分别乘以其相对于偏好序列给定的数量，根据定义式(7.3)，分别计算出各因子的模糊相对权重，即

$$\sum_{l=1}^{5} l \cdot \mu_{形状,l} = 1\times(0.10)+2\times(0.35)+3\times(0.50)+4\times(0.75)+5\times(0.65)=8.55$$

$$\sum_{l=1}^{5} l \cdot \mu_{色泽,l} = 1\times(0.30)+2\times(0.50)+3\times(0.60)+4\times(0.20)+5\times(0.10)=4.4$$

$$\sum_{l=1}^{5} l \cdot \mu_{水色,l} = 1\times(0.20)+2\times(0.45)+3\times(0.80)+4\times(0.50)+5\times(0.30)=7$$

$$\sum_{l=1}^{5} l \cdot \mu_{香味,l} = 1\times(0)+2\times(0.10)+3\times(0.40)+4\times(0.80)+5\times(0.92)=9.2$$

因为 $\sum_{i=1}^{k}\sum_{l=1}^{f} l\times\mu_{S_i,l} = 8.55+4.4+7+9.2=29.15$，所以这四个项目的模糊相对权重分别为

$$FRW_{形状} = \frac{8.55}{29.15} \approx 0.29$$

$$FRW_{色泽} = \frac{4.4}{29.15} \approx 0.15$$

$$FRW_{水色} = \frac{7}{29.15} \approx 0.24$$

$$FRW_{香味} = \frac{9.2}{29.15} \approx 0.32$$

故得形状因子的权重为0.29，色泽因子的权重为0.15，水色因子的权重为0.24和香味因子的权重为0.32。

7.3 加权模糊分类

所谓加权模糊聚类分析法，即加权模糊 c 平均聚类法（Weighted Fuzzy c Mean Clustering Method，WFCM）。设有 n 个样本 $x_1, x_2, \cdots, x_n$，每个样本有 f 种特性指标，用 x_{jl} 表示第 j 个样本第 l 个指标的观测值；而用 $\bar{x}_l$ 与 $\hat{\sigma}_l$ 表示第 l 个特性指标的样本平均值与样本标准偏差。需要特别强调的是，如果评估项目的单位不同，为了除去评判依据的差异，必须先对原始数据作标准化处理。我们考虑用

$$x_{jl}^* = \frac{x_{jl} - \bar{x}_l}{\hat{\sigma}_l}; j = 1,2,\cdots,n; l = 1,2,\cdots,f \tag{7.11}$$

作为标准化后的观测值数据；反之，如果评估项目的单位相同，为了避免原始数据的变换或变形，则不需另作此项处理，以保留最初数据的信息。

令 $P = \{P_i, i = 1,2,\cdots,c\}$ 为论域 X 上的一个模糊分割。根据模糊集的概念，每一样本点可以有属于一个以上类的隶属度。因而，定义第 j 个数据点属于第 i 类的隶属度为 $\mu_{ij} = \mu_{P_i}(x_j) \in [0,1]$ 且满足以下两个约束式

$$\sum_{i=1}^{c} \mu_{ij} = 1, j = 1,2,\cdots,n \tag{7.12}$$

$$0 < \sum_{j=1}^{n} \mu_{ij} < n \tag{7.13}$$

对 c 类分类与 n 个数据点定义一个加权模糊 c 分割空间，也就是 M_{wfc}

$$M_{wfc} = \{P \mid \mu_{ij} \in [0,1], \sum_{i=1}^{c} \mu_{ij} = 1, 0 < \sum_{j=1}^{n} \mu_{ij} < n\} \tag{7.14}$$

故对于任何的 $P \in M_{wfc}$，一个加权模糊分割矩阵 P。

本节所介绍的加权模糊聚类分析法是以最小误差平方标准为基础的聚类算法。假设代表各样本的统计特性指标因子的模糊相对权重为 $FRW = (FRW_1, FRW_2, \cdots, FRW_f)$，则加权模糊分割矩阵 P 的聚类准则为

$$J_{wfc}(P, v) = \sum_{j=1}^{n} \sum_{i=1}^{c} (\mu_{ij})^m (d_{ij})^2 \tag{7.15}$$

其中 $d_{ij} = d(x_j \cdot FRW - v_i) = \left[\sum_{l=1}^{f} (x_{jl}^{(t)} FRW_l - v_i^{(t)})^2\right]^{1/2}$，第 i 个类中心到第 j 个数据点的欧几里得距离，μ_{ij} 表示第 j 个数据点属于第 i 类的隶属度，v_i 表示第 i 个类中心，FRW_l 表示第 l 个评价因素的模糊相对权重值。如果考虑个别时期的分类，只要对第 t 期聚类准则条件下求 μ_{ij} 及 v_i 使满足

$$\min J_{wfc}^{(t)}(P, v^{(t)}) = \sum_{j=1}^{n} \sum_{i=1}^{c} (\mu_{ij})^m (d_{ij}^{(t)})^2 \tag{7.16}$$

$$\text{s.t}\quad \sum_{i=1}^{c}\mu_{ij}=1,0<\sum_{j=1}^{n}\mu_{ij}<n \tag{7.17}$$

利用拉格朗日乘数法(Lagrange Multiplier Method,又称为拉格朗日乘子法),将式(7.16)分别对 μ_{ij} 及 v_i 求偏微分后,得到第 j 个样本对第 i 类类中心的隶属度

$$u_{ik}=\left[\sum_{i=1}^{c}\left\{\frac{\sum_{j=1}^{m}(x_{ij}^{(t)}FRW_t-v_{ij}^{(t)})^2}{\sum_{l=1}^{f}(x_{ij}^{(t)}FRW_t v_{ij}^{(t)})^2}\right\}^{\frac{1}{m-1}}\right]^{-1} \tag{7.18}$$

及第 i 类的类中心

$$v_i^{(t)}=\frac{FRW_t\sum_{j=1}^{n}(\mu_{ij})^m x_{ji}^{(t)}}{\sum_{j=1}^{n}(\mu_{ij})^m} \tag{7.19}$$

上式解并不能保证为全局的最理想解。但是,我们要求在事前指定的精确水平范围内,所求得的最佳有效解。

模糊统计分类评定的步骤如下:

步骤 1:定义想要分类样本其评估项目的模糊相对权重 FRW。

步骤 2:选择 c 个初始类中心,初始模糊分类 $P^{(0)}$, m 以及误差值 ϵ。

步骤 3:由式(7.19)计算 $P^{(t)}$ 的 c 个类中心 $v^{(t)}=\{v_1^{(t)},v_2^{(t)},\cdots,v_c^{(t)}\}$。

步骤 4:对第 t 步骤,更新分割矩阵 $P^{(t)}=(\mu_{ij}^{(t)})$,计算新的模糊分类

$$P^{(t+1)}=\{P_1^{(t+1)},P_2^{(t+1)},\cdots,P_c^{(t+1)}\}$$

步骤 5:比较 $P^{(t)}$ 与 $P^{(t+1)}$,若 $\|P^{(t)}-P^{(t+1)}\|<\epsilon$,运算就停止。否则,令 $t=t+1$,回到步骤 2。

步骤 6(硬分类):对样本点至各类中心的隶属度中选取最大值,即为该点隶属于该类。

[评注]

聚类分析是相当实用的统计技术,虽然一些学者已提出不少方法,但是对下面几点仍要加以注意,其中的一些问题未来仍有研究的空间:

(1) 权重的重要性:一般的分类法忽视了各因子间重要程度的不同,如此势必影响到分类的结果,造成不当的推断。而权重的考虑及决定权数更需要详加探讨,这也是本章采用加权模糊分类法的原因。

(2) 隶属度的分布:传统的分类法,把所有事物看做具有清楚、不含糊的界限,将每一个辨识对象严格地划分为属于某类,并认为元素 x 不是属于 A 集合,就是不属于 A 集合,即特征函数的观念。但是,在实际生活中,对某些样本来说

并不具有严格的属性，它们可能位于两种聚类之间，这时采用模糊分类可能会获得更好的结果。换句话说，对应于人脑的思维所发展出来的自然语言，并不是非此即彼而是亦此亦彼的观念。

(3) 时间因素的考虑：一般地说，人们在时间 t 的分类结果与时间 $t+1$ 的分类结果不一定相同，如此也会造成不同时期分类的不一致。一般的分类法，并没有考虑时间因素，或对于不同时期的评分多取其平均法；若考虑将时间的权重因素加入进来作聚类分析，使数据形成动态的模式，更能充分地表现整体的分类结果。

7.4 茶叶等级分类实例

茶叶的种类繁多，不同消费群体及环境所要求的质量或喜好的程度不同，以致茶叶的价值难以评估及量化。通常，茶叶的质量评定方式是依据人的感觉功能及其经验为主，在最短时间内判断各种茶叶的外观和汤质。而茶叶的质量受到品种、栽培管理、气候环境以及制茶技术等因素所影响，加上每个评茶人员主观喜好、表达方式不同，造成评分标准的差异，这无疑相对增加了茶叶分类的困难度。

对茶叶质量进行分析与鉴定不仅是为了分级或比赛，对于品种改良、栽培、制茶与贮存方法等都有很大的影响。茶的好坏，有赖于正确的茶叶分级制度。因此，茶叶质量的鉴定与分类是一件非常重要的工作。本节将以中国台湾 69 个茶叶数据，应用不同的分类方法进行分析，并比较不同分类方法在应用上的差异性。

我们以中国台湾某茶业改良场品种园栽植的白毛猴等 69 个茶树品种，分别对其春、夏、秋三季及全年制茶质量进行茶叶官能质量评估，探讨茶叶品种间的质量差异。

评分

茶叶优良与否乃是模糊概念，其分级方式通常是根据外观与汤质。茶叶的外观分为形状和色泽二项目，而汤质可分为水色及香味二项目。因此，我们选择评判茶叶质量的四个评估项目为形状、色泽、水色及香味。

一般标准评茶步骤如下：随机秤取茶样 3 g，倒入审茶杯中，注入沸水 142 ml，加盖静置 5 ~ 6 min，倒出茶汤于茶碗中，再分别依形状、色泽、水色及香味等四项目成绩评审。

本研究邀请 10 位评茶专家对每种茶叶进行茶叶质量的鉴定。将形状、色泽、水色及香味四方面直接评分。采取 25 分制，满分为 100 分。并把每位评分员

对某一茶叶，某一项目的评分进行算术平均，进而以评定加权总分数的高低作为决定茶叶品级的标准。

因为形状、色泽、水色及香味此四项目均以分作为评估的单位，因此，评判依据的差异性不大，故不需作标准化处理。设第 K 个评分员（$K = 1,\cdots,10$）对某一茶叶、某一因素的评分为 $x_{ij(k)}$（$i = 1,2,\cdots,69, j = 1,2,3,4$），那么 10 个评分员对某一茶叶的某一项目评分平均值为

$$\bar{x}_{ij} = \frac{1}{10}\sum_{k=1}^{10} x_{ij(k)};i = 1,2,\cdots,69;j = 1,2,\cdots,4$$

最后，得到 69 种茶叶依春、夏、秋三季及全年的最终评分数据。因篇幅关系，我们仅列春季及全年的数据。

因子加权

因子加权，也就是对四个因素在评价茶叶时的重要性给予赋权。茶叶原有一股香味，即使是同样的茶，但因其品级不同而风味大有差别，而且中国人又比其他国家的人民更重视茶叶的香味。因此，显然香味的权数较其他因素的权数来得大。另外，茶叶的质量又因其采收季节的不同而有所差异；一般地说，春茶在相同的茶叶中是较优良的茶叶。所以，在进行茶质鉴定及分析时，应该加入权重的考虑，以期反映真正分类结果。

因此，请 10 位评茶专家利用模糊问卷法，对茶叶的评估项目及季节进行重要程度的评定，得到表 7.4 及表 7.5 的平均后模糊评估表。

表 7.4　制茶质量评估项目隶属度的比较

评估项目	很不重要	不重要	普通	重要	很重要
形　状	0.50	0.60	0.30	0.50	0.20
色　泽	0.30	0.40	0.47	0.60	0.30
水　色	0.60	0.57	0.40	0.30	0.10
香　味	1.0	0.80	0.50	0.20	0.07

表 7.5　制茶质量季节性评估隶属度的比较

评估项目	很不重要	不重要	普通	重要	很重要
春　茶	0.80	0.50	0.50	0.37	0.10
夏　茶	0.30	0.60	0.80	0.50	0.40
秋　茶	0.40	0.50	0.57	0.60	0.30

当决定形状、色泽、水色及香味等评估项目和春茶、夏茶及秋茶等季节的模糊权重后，再求出其模糊相对权重值。由此可得，评估项目（形状，色泽，水色，香味）的模糊相对权重为 $W = (w_1, w_2, w_3, w_4) = (0.23, 0.19, 0.25, 0.33)$，与季节（春季，夏季，秋季）的模糊相对权重 $S = (s_1, s_2, s_3) = (0.37, 0.31, 0.32)$。

分类结果分析与比较

根据茶叶专家的建议，茶叶的分类多倾向于分成五类。即由最好到最差分别为第一级到第五级。我们用以下 4 种方法，分别将 69 种茶树品种制茶质量按春茶、夏茶、秋茶及全年，依据不同时期分成 7 类（也就是聚类）。（参看附表 7.1 及 7.2）

a. 传统分类法：利用聚类分析中的 K 平均法对 69 个制茶质量分类。

b. 类神经网络分类法：利用神经网络中逆传导训练法对茶叶作分类，先以传统分类结果的五组聚类中心去作网络学习二万次或 $\epsilon < 0.01$，再将 69 个茶叶进入神经网络作分类。

c. 模糊分类法：利用模糊 c 平均法作茶叶分类。为了使分类结果较为保守，故拟定加权指数 $m = 2$ 且取 $\epsilon_L = 10^{-6} > 0$ 为停止迭代标准。

d. 加权模糊分类法：模糊权重的重要信息对分类影响甚大，故加入茶叶评估项目及季节的模糊相对权重。以加权指数 $m = 2$ 且取 $\epsilon_L = 10^{-6} > 0$ 为停止迭代标准及模糊问卷所得的权重，对 69 个品种制茶质量进行分类。

从分类结果来看，茶叶等级以春茶为佳，秋茶次之，夏茶最次。此结果与传统茶叶相似。而在茶叶等级评定方面，则有些不同。这里将季节间茶叶质量分类结果分析如下。

春茶品质评定

a. 白毛猴、黑毛猴及贵仔坑白毛猴，由于形状自然卷曲，色泽墨绿，滋味甘醇，因此，不论任何分类法，其质量归类于第一级。

b. 经模糊分类后，中国台湾四大品种中青心乌龙及青心大有归类于第一级，大叶乌龙及硬枝红心归类于第三级；而经加权模糊分类后，香味质量有平均表现的大叶乌龙种则归类提高至第二级。

c. 模糊分类隶属第二级的红心乌龙、福州种、铁观音、横这大叶、桂花种、牛埔种、黑面早种、青心早种、汉口种、平水种、大吉岭种等 11 个品种，经模糊加权分类后则提升至第一级。

d. 大叶乌系的林口大叶乌种及文山大叶乌种，不论模糊加权与否，仍归类于第三级，显示形状、色泽等评估项目平均表现。

e.适制红茶的阿萨姆、山茶、Kyang、Shan及Manipuri等五个品种,经模糊分类后归属于第五级,由于香气低、苦涩味重,经模糊加权分类后仍属第五级。

夏茶品质评定

a.春茶不论任何分类法均归属第一级的白毛猴、黑毛猴及贵仔坑白毛猴,由于夏茶制茶质量较为接近,经模糊加权分类后与春茶属第三级的大南湾白毛猴融合为第一级。

b.春茶依模糊加权分类后隶属于第一级的平水种、大吉岭种,第二级的基隆金龟、大叶乌龙、小叶竹叶、伸蔓种,第三级的牛屎乌、盐川种、黄柑、枝兰种、竹叶种、白种、宇治种,与第四级的莳茶、白心武夷、基隆白种、黄茶,因其夏茶的评估项目表现平均,故不论模糊加权与否均归类于第四级。

c.奉卢种形状粗松、水色暗黄、滋味苦涩,而与阿萨姆、Kyang等适制红茶品种融合为一群,归类于第五级。

秋茶品质评定

a.传统分类法将乌龙系的青心乌龙、红心乌龙、白心乌龙与大叶乌龙归类为第二级,经模糊分类后,青心乌龙、红心乌龙与白心乌龙提升至第一级,大叶乌龙归类至第三级;而加权模糊分类后则青心乌龙、红心乌龙与大叶乌龙种归类为第三级,白心乌龙提升至第一级。

b.形状、色泽及香味等项目评分相同的文山大叶乌及林口大叶乌品种,不论任何分类法均融合为同群,显示秋茶质量评定相近归类为同等级茶。

c.形状及香味评分相同的横这大叶、青心早种、湖南种、金龟种、猫耳种、基隆白种、小叶竹叶及白种,依据传统分类、神经网络及模糊分类法均同属第四级,而经加权模糊后则提升至第二级,显示香味有较突出的表现。

d.福州种、桃仁种、柑仔种、三叉枝兰、水仙、桂花种、柑仔种(黄心)、埔心种、黄枝种及大湖尾种,因质量接近而不论任何分类法均归属第三级。

全年茶种品质评定

a.白毛猴系的白毛猴及大南湾白毛猴种,因其制茶质量优良,全年皆属第一级。

b.青心大有、青心乌龙、硬枝红心及大叶乌龙等台湾四大品种,全年制茶质量经加权模糊后青心大有与青心乌龙种归属第一级,硬枝红心与大叶乌龙种归属第三级。

c.竹叶系的小叶竹叶及竹叶种,其全年制茶品质同属第四级外,适制包种

茶之春秋季节其制茶质量有差异，而文山大叶乌及林口大叶乌种依传统分类法或神经网络分类法均归类为第三级，而经模糊分类后则归类至第二级。

7.5 结　　论

社会科学的问题研究极其复杂且多样化，传统的分类方法已经无法满足各种不同的理论及观点。而传统的分类法大都属于硬分类，即分类对象完全属于某群集。这是其不合常理之处。又因为分类对象所考虑的论域真子集，有边界不清、权重模糊的性质，故有采用模糊加权分类法的必要。

本章我们比较了传统分类与模糊分类的技术，并发展出一套新的模糊分类方法，即加权的模糊 c 分类法。更进一步，把这种方法应用在茶叶的分级上。以模糊统计法来估计分类论域中各因子的模糊权重及模糊相对权重。由此，对茶叶质量评定应用作不同分类法的分析比较。经过分类结果可知，就茶叶质量评定而言，以加权模糊分类法的结果最为符合实际情况的分类。如何将隶属度概念，将模糊分类方法应用于一般实际分类评估中，更进一步地用模糊统计方法作为数据挖掘的依据，则是未来发展非常具有前景的研究领域。

附表 7.1　中国台湾 69 个茶树品种春季制茶质量的评分比较

茶叶品种	形状	色泽	水色	香味	合计	名次	茶叶品种	形状	色泽	水色	香味	合计	名次
白毛猴	22.3	15.5	15.7	21.0	74.5	2	牛屎乌	20.3	13.3	15.2	18.3	67.1	39
黑毛猴	22.0	15.4	16.5	21.4	75.3	1	桃仁乌	20.3	12.8	14.8	18.3	66.2	45
青心大有	21.5	15.3	16.0	20.0	72.8	5	湖南种	20.2	13.5	14.5	19.0	67.2	38
贵仔坑白毛猴	22.0	15.5	16.3	20.5	74.3	3	黄枝种	19.0	12.3	14.5	19.5	65.3	51
大南湾白毛猴	21.0	12.0	13.0	19.0	65.0	52	大湖尾	20.3	13.6	14.3	19.4	67.6	34
青心乌龙	21.0	14.8	16.0	21.5	73.3	4	基隆金龟	20.0	13.8	14.8	19.2	67.8	33
红心乌龙	20.0	13.5	14.3	21.0	68.8	28	平水种	20.1	13.5	16.1	20.5	70.2	15
福州种	20.8	14.2	14.8	21.2	71.0	9	盐川种	20.4	13.1	14.6	18.2	66.3	44
淡水青心	20.9	14.0	16.4	20.6	71.9	7	乌金种	19.0	13.0	12.5	18.0	62.5	59
铁观音	19.0	14.0	15.0	21.0	69.0	26	金龟种	19.5	13.0	13.3	17.5	63.3	58
不知春	21.0	14.3	15.5	20.5	71.3	8	大吉岭种	19.8	13.1	16.1	20.5	69.5	18
桃仁种	21.3	14.4	16.1	21.0	72.8	5	黄柑	19.8	13.6	14.4	18.9	66.7	42
晚种	20.7	14.3	15.2	19.3	69.5	18	枝兰种	19.3	12.6	15.0	19.3	66.2	45
红心大有	20.6	14.3	15.4	19.2	69.5	18	莳茶	20.6	13.7	15.0	17.2	66.5	43
横这大叶	20.0	13.9	15.4	21.3	70.6	12	白心武夷	20.0	13.5	13.5	18.0	65.0	52
白心乌龙	20.7	13.6	14.3	20.1	68.7	29	猫耳种	20.0	13.0	14.0	19.0	66.0	48
柑仔种	21.3	13.8	14.8	20.2	70.1	16	基隆白种	20.7	12.2	13.7	17.0	63.6	57
三叉枝兰	20.5	13.4	13.4	18.8	66.1	47	小叶竹叶	20.0	13.3	15.0	19.3	67.6	34
硬枝红心	20.3	14.0	13.7	18.8	66.8	41	伸蔓种	19.5	13.0	15.0	19.5	67.0	40
水仙	20.5	14.0	14.8	20.0	69.3	22	竹叶种	20.5	13.0	13.0	18.5	65.0	52
桂花种	19.3	13.7	15.5	22.0	70.5	13	白种	18.5	13.0	16.0	18.0	65.5	50
大叶乌龙	20.5	14.8	14.3	19.5	69.1	25	白叶种	19.8	12.1	14.4	18.3	64.6	55
牛埔种	20.3	13.9	16.3	20.0	70.5	13	宇治种	21.0	12.6	13.4	18.7	65.7	49
天公种	21.5	13.8	14.4	19.9	69.6	17	黄茶	19.0	10.0	14.0	18.0	61.0	60
柑仔种(黄心)	21.3	12.5	13.8	20.0	67.6	34	秦卢种	19.1	12.1	12.6	16.8	60.6	61
文山大叶乌	20.5	14.5	15.2	18.7	68.9	27	Manipuri	21.0	11.8	11.8	11.0	55.6	66
林口横这	20.6	13.7	14.6	19.5	68.4	30	Indigenous	20.3	11.5	12.5	14.5	58.8	62
黑面早种	20.6	13.6	15.2	21.5	70.9	10	南投山茶	20.0	11.8	12.3	8.7	52.8	69
青心早种	19.8	13.7	15.6	21.8	70.9	10	Japuri	21.0	11.0	11.5	11.5	55.0	67
枫仔林	21.6	13.2	14.0	19.5	68.3	31	阿萨姆	19.5	11.0	13.0	11.5	55.0	67
林口大叶乌	20.5	13.4	15.0	18.6	67.5	37	缅甸种	21.0	14.0	13.5	16.0	64.5	56
早种	20.4	14.0	15.3	19.6	69.3	22	山茶	20.3	11.0	12.5	12.7	56.5	65
大叶竹叶	20.4	13.3	15.1	20.4	69.2	24	Kyang	20.3	12.3	12.0	12.7	57.3	64
埔心种	20.5	13.6	14.6	19.5	68.2	32	Shan	21.0	10.6	12.2	14.3	58.1	63
汉口种	19.7	13.5	16.0	20.3	69.5	18							

附表 7.2 中国台湾 69 个茶树品种全年制茶质量的评分比较

茶叶品种	形状	色泽	水色	香味	合计	名次	茶叶品种	形状	色泽	水色	香味	合计	名次
白毛猴	23.5	15.7	15.3	21.9	76.4	1	牛屎乌	20.7	14.0	15.2	18.8	68.7	35
黑毛猴	22.0	15.1	15.7	21.9	74.7	2	桃仁乌	20.2	13.2	14.7	19.5	67.6	45
青心大冇	21.2	14.9	15.3	22.1	73.5	3	湖南种	20.2	13.7	14.3	19.6	67.7	44
贵仔坑白毛猴	22.1	15.1	15.0	20.9	73.2	5	黄枝种	19.8	13.2	14.1	20.8	68.0	41
大南湾白毛猴	22.5	14.5	14.3	20.8	72.2	7	大湖尾	19.9	13.4	14.9	20.6	68.7	34
青心乌龙	20.9	14.8	15.7	22.0	73.3	4	基隆金龟	20.2	14.1	14.7	19.3	68.3	37
红心乌龙	21.0	14.5	14.8	22.0	72.3	6	平水种	19.7	13.3	15.1	19.9	68.1	39
福州种	20.7	14.0	14.8	22.2	71.8	8	盐川种	20.3	13.6	15.0	19.7	68.6	36
淡水青心	20.9	14.1	15.3	21.0	71.3	9	乌金种	19.7	13.7	14.2	20.0	67.5	48
铁观音	19.7	13.8	14.4	22.4	70.4	18	金龟种	20.0	13.5	13.9	19.3	66.8	50
不知春	21.0	14.1	15.1	21.0	71.3	9	大吉岭种	19.6	13.2	15.1	19.8	67.6	45
桃仁种	20.7	13.7	15.3	21.1	70.8	12	黄柑	19.9	13.7	14.8	19.3	67.8	43
晚种	20.9	14.4	15.0	20.3	70.6	13	枝兰种	20.1	13.2	14.8	19.4	67.6	45
红心大冇	20.4	14.4	15.5	20.8	71.1	11	莳茶	20.9	14.1	14.8	18.3	68.1	40
横这大叶	20.5	13.9	14.7	20.7	69.8	25	白心武夷	20.1	13.5	14.0	19.1	66.7	51
白心乌龙	21.2	14.3	14.9	20.2	70.6	14	猫耳种	20.2	13.3	14.1	19.5	67.1	49
柑仔种	20.7	14.2	14.5	20.9	70.3	19	基隆白种	20.3	13.0	14.4	18.4	66.1	53
三叉枝兰	20.7	14.2	14.3	20.6	69.8	23	小叶竹叶	19.1	13.3	14.6	19.0	66.1	53
硬枝红心	20.6	14.4	14.4	20.4	69.8	26	伸蔓种	19.3	13.1	14.6	19.3	66.2	52
水仙	20.0	13.8	15.1	21.2	70.1	21	竹叶种	20.5	13.3	13.7	18.5	65.9	55
桂花种	19.9	13.9	15.2	21.5	70.5	16	白种	19.4	13.1	14.8	18.7	65.9	55
大叶乌龙	20.8	14.7	14.6	20.5	70.6	15	白叶种	19.4	12.1	14.3	19.5	65.3	57
牛埔种	20.5	13.9	15.3	20.6	70.2	20	宇治种	20.0	13.0	13.4	17.9	64.2	58
天公种	21.2	14.2	14.7	20.3	70.5	17	黄茶	19.6	12.2	13.7	17.9	63.4	59
柑仔种(黄心)	20.9	13.5	14.4	20.8	69.5	30	皋卢种	18.9	11.8	13.4	13.5	57.7	61
文山大叶乌	20.5	14.6	15.0	19.8	69.9	22	Manipuri	20.1	12.3	12.8	12.4	57.6	62
林口横这	20.5	13.7	14.9	20.6	69.6	28	Indigenous	19.2	12.0	13.5	13.9	58.5	60
黑面早种	20.5	13.9	14.8	20.4	69.6	28	南投山茶	19.6	12.7	13.2	11.1	56.6	65
青心早种	19.1	13.3	15.0	20.5	67.9	42	Japuri	19.8	11.7	12.7	12.4	56.6	65
枫仔林	20.6	13.6	14.9	20.3	69.4	31	阿萨姆	19.3	11.8	13.8	11.9	56.7	64
林口大叶乌	20.9	14.1	14.7	20.0	69.7	27	缅甸种	19.8	12.1	12.9	12.3	57.2	63
早种	20.4	14.1	14.8	20.5	69.8	24	山茶	19.6	11.6	13.5	11.3	55.9	68
大叶竹叶	19.8	13.4	14.8	21.0	69.0	33	Kyang	19.3	12.2	12.9	11.6	56.0	67
埔心种	20.2	13.8	14.9	20.5	69.4	32	Shan	20.1	11.1	12.1	11.8	55.0	69
汉口种	19.4	13.4	15.3	20.0	68.1	38							

附表 7.3　中国台湾 69 个茶树品种春茶的分类结果

茶叶品种	传统分类	神经网络	模糊分类	加权模糊	茶叶品种	传统分类	神经网络	模糊分类	加权模糊
白毛猴	1	1	1	1	牛屎乌	4	3	3	3
黑毛猴	1	1	1	1	桃仁乌	4	4	3	3
青心大有	2	1	1	1	湖南种	4	3	3	3
贵仔坑白毛猴	1	1	1	1	黄枝种	4	4	3	3
大南湾白毛猴	4	4	4	3	大湖尾	3	3	3	2
青心乌龙	2	2	1	1	基隆金龟	3	3	3	2
红心乌龙	3	3	2	1	平水种	3	3	2	1
福州种	3	3	2	1	盐川种	4	4	3	3
淡水青心	2	2	1	1	乌金种	4	4	4	4
铁观音	3	3	2	1	金龟种	4	4	4	4
不知春	2	2	1	1	大吉岭种	3	3	2	1
桃仁种	2	2	1	1	黄柑	4	4	3	3
晚种	3	3	3	2	枝兰种	4	4	3	3
红心大有	3	3	3	2	药茶	4	3	4	4
横这大叶	3	3	2	1	白心武夷	4	4	4	4
白心乌龙	3	3	3	2	猫耳种	4	4	3	3
柑仔种	3	2	2	2	基隆白种	4	4	4	4
三叉枝兰	4	4	4	3	小叶竹叶	3	3	3	2
硬枝红心	4	4	3	3	伸蔓种	3	4	3	2
水仙	3	3	2	2	竹叶种	4	4	4	3
桂花种	3	3	2	1	白种	4	4	3	3
大叶乌龙	3	3	3	2	白叶种	4	4	4	3
牛埔种	3	3	2	1	宇治种	4	4	4	3
天公种	3	2	3	2	黄茶	4	4	4	4
柑仔种(黄心)	4	3	3	2	摹卢种	4	5	4	4
文山大叶乌	3	3	3	3	Manipuri	5	5	5	5
林口横这	3	3	3	2	Indigenous	5	5	5	4
黑面早种	3	3	2	1	南投山茶	5	5	5	5
青心早种	3	3	2	1	Japuri	5	5	5	5
枫仔林	4	2	3	2	阿萨姆	5	5	5	5
林口大叶乌	4	3	3	3	缅甸种	4	3	4	4
早种	3	3	3	2	山茶	5	5	5	5
大叶竹叶	3	3	2	2	Kyang	5	5	5	5
埔心种	3	3	3	2	Shan	5	5	5	5
汉口种	3	3	2	1					

附表 7.4 中国台湾 69 个茶树品种全年的分类结果

茶叶品种	传统分类	神经网络	模糊分类	加权模糊	茶叶品种	传统分类	神经网络	模糊分类	加权模糊
白毛猴	1	1	1	1	牛屎乌	4	3	3	4
黑毛猴	2	1	1	1	桃仁乌	4	4	3	2
青心大有	2	2	1	1	湖南种	4	4	4	4
贵仔坑白毛猴	2	2	1	1	黄枝种	3	4	2	3
大南湾白毛猴	2	2	1	1	大湖尾	3	4	2	3
青心乌龙	2	2	1	1	基隆金龟	4	3	3	2
红心乌龙	2	2	1	1	平水种	4	4	3	2
福州种	2	3	1	1	盐川种	4	3	2	3
淡水青心	3	2	1	3	乌金种	4	4	2	3
铁观音	3	4	1	1	金龟种	4	4	4	4
不知春	3	2	1	3	大吉岭种	4	4	3	2
桃仁种	3	3	2	3	黄柑	4	4	3	4
晚种	3	2	2	3	枝兰种	4	4	3	2
红心大有	3	3	2	3	附茶	4	3	3	4
横这大叶	3	3	3	2	白心武夷	4	4	4	4
白心乌龙	3	2	2	3	猫耳种	4	4	4	4
柑仔种	3	3	2	3	基隆白种	4	4	4	4
三叉枝兰	3	3	2	3	小叶竹叶	4	4	4	4
硬枝红心	3	3	2	3	伸蔓种	4	4	4	4
水仙	3	3	2	3	竹叶种	4	4	4	4
桂花种	3	3	2	3	白种	4	4	4	4
大叶乌龙	3	3	2	3	白叶种	4	4	4	4
牛埔种	3	3	2	3	宇治种	4	4	4	4
天公种	3	2	2	3	黄茶	4	4	4	4
柑仔种(黄心)	3	3	2	3	皋卢种	5	5	5	5
文山大叶乌	3	3	2	2	Manipuri	5	5	5	5
林口横这	3	3	2	3	Indigenous	5	5	5	5
黑面早种	3	3	3	2	南投山茶	5	5	5	5
青心早种	3	4	4	2	Japuri	5	5	5	5
枫仔林	3	3	2	3	阿萨姆	5	5	5	5
林口大叶乌	3	3	2	3	缅甸种	5	5	5	5
早种	3	3	2	3	山茶	5	5	5	5
大叶竹叶	3	4	2	3	Kyang	5	5	5	5
埔心种	3	3	2	3	Shan	5	5	5	5
汉口种	4	4	3	2					

附表 7.5　中国台湾 69 个茶树品种春茶制茶质量加权模糊聚类分析

聚类类别	茶叶品种
第一级	白毛猴、黑毛猴、青心大有、贵仔坑白毛猴、青心乌龙、红心乌龙、福州种、淡水青心、铁观音、不知春、桃仁种、横这大叶、桂花种、牛埔种、黑面早种、青心早种、汉口种、平水种、大吉岭种
第二级	晚种、红心大有、白心乌龙、柑仔种、水仙、大叶乌龙、天公种、柑仔种(黄心)、林口横这、枫仔林、早种、大叶竹叶、埔心种、大湖尾、基隆金龟、小叶竹叶、伸蔓种
第三级	大南湾白毛猴、三叉枝兰、硬枝红心、文山大叶乌、林口大叶乌、牛屎乌、桃仁乌、湖南种、黄枝种、盐川种、黄柑、枝兰种、猫耳种、竹叶种、白种、白叶种、宇治种
第四级	乌金种、金龟种、莳茶、白心武夷、基隆白种、黄茶、阜卢种、Indigenous、缅甸种
第五级	Manipuri、南投山茶、Japuri、阿萨姆、山茶、Kyang、Shan

附表 7.6　中国台湾 69 个茶树品种全年制茶质量加权模糊聚类分析

聚类类别	茶叶品种
第一级	白毛猴、黑毛猴、青心大有、贵仔坑白毛猴、大南湾白毛猴、青心乌龙、红心乌龙、福州种、铁观音
第二级	横这大叶、文山大叶乌、黑面早种、青心早种、汉口种、桃仁乌、基隆金龟、平水种、大吉岭种、枝兰种
第三级	淡水青心、不知春、桃仁种、晚种、红心大有、白心乌龙、柑仔种、三叉枝兰、硬枝红心、水仙、桂花种、大叶乌龙、牛埔种、天公种、柑仔种(黄心)、林口横这、枫仔林、林口大叶乌、早种、大叶竹叶、埔心种、黄枝种、大湖尾、盐川种、乌金种
第四级	牛屎乌、湖南种、金龟种、黄柑、莳茶、白心武夷、猫耳种、基隆白种、小叶竹叶、伸蔓种、竹叶种、白种、白叶种、宇治种、黄茶
第五级	阜卢种、Manipuri、Indigenous、南投山茶、Japuri、阿萨姆、缅甸种、山茶、Kyang、Shan

第 8 章　模糊数据回归模型及应用

传统的统计回归模型假定观测值的不确定性源自随机现象，模糊数据回归则考虑不确定性源自多重隶属现象。

模糊数据回归利用样本模糊数(x_i, Y_i)来对模糊数据回归参数进行估计，其中 Y_i 表示观测模糊数。本章将讨论估计模糊参数的二种方法：(1) 线性规划法，此方法将求出适当的区间，来把观测模糊数 Y_i 的分布范围全部覆盖住；(2) 最小二乘法，该方法运用统计误差分布的观点，找出模糊数 Y_i 的集中趋势区间。

回归模型经常用来建立金融与经济等方面的模型，而此种模型常常带有模糊特性，例如，景气循环、不规则趋势等。最后，针对中国台湾景气指标进行实证分析研究。

8.1　模糊数据回归简介

我们经常使用回归来描述自变量 X 和因变量 Y 之间的因果关系。回归参数的估计是回归分析的主要研究课题。传统回归分析均假定，观测值是由实际因变量影响加上随机误差而产生的。也就是说，因变量是一种带有不确定性的随机变量。然而，在许多实际应用中，往往无法满足这样的假设。例如，有些因变量 Y 的观测值并不是以单一数值的形式存在，而是带有多重隶属的特性。这些观测值虽然带有不确定性，但这种不确定的特性是来自于多重隶属现象，而非随机现象。

以 2007 年 6 月 1 日人民币对美元的汇率中间价为例，统计数字为人民币 / 美元 = 7.649 7，而 6 月 29 日人民币对美元的汇率中间价为 7.615 5。但是，这数据只是交易日的数值数据，它并不能完全准确地反映出在 2007 年 6 月 1 日与 6 月 29 日两个时间区间里，人民币对美元汇率的变动状况。因此，人民币对美元汇率这个数字的本身就具有不确定性与模糊性。在这些情况下，如果利用假设的“精确值”，可能会误导模型的建立，也可能扩大预测结果和实际状态之间的误差。

近年来，一些学者认为，在人文社会科学的测度理念里模糊统计和模糊相

关性的使用是对数值模型的推广。模糊统计和模糊相关性日渐受到重视，这顺应了复杂的人文社会现象无法用传统模型解释而发展的需求。传统的数字数据有需求过度及过度解释的危险，如果应用模糊的数值数据，可以比较好地避免这类危险的发生。

为了理清、改变传统对统计数据的"随机"观点，现在以另一种视角来看待不确定性的问题，我们考虑用模糊观点来取代随机观点。倘若运用模糊观点套用回归分析来处理模糊数据问题，就称为模糊数据回归分析。模糊数据回归分析最早是由田中、上岛和浅井(Tanaka, Uejima and Asai, 1980)提出的。一般模糊数据回归能够处理的样本形式有两种：一种是实数值样本(x_i, y_i)，其中 x_i 表示实数向量，y_i 表示观测到的实数值；一种是模糊数样本(x_i, Y_i)，其中 Y_i 不是实数形式，而是观测到的模糊数。之所以可采用实数值样本(x_i, y_i)来推断模糊数据回归模型，这是因为人们相信观测值(x_i, y_i)内含有强烈的隶属度信息。

然而，这样的说法亦缺乏说服力，因为隶属度是分布在区间上的，而样本(x_i, y_i)则是以实数值的形式存在，在实数值上能否蕴含正确的隶属度信息是值得商榷的。要得到正确的估计结果，最好是收集具有代表性的隶属函数样本，所谓具有代表性的隶属函数样本应该是以模糊数的形式存在。因而，我们比较支持用样本模糊数(x_i, Y_i)来推断模糊数据回归模型，本章将讨论样本为模糊数的情况。

8.2　模糊数据回归建构

线性模糊数据回归表示成

$$Y(x_i) = A_0 + A_1x_{1i} + A_2x_{2i} + \cdots + A_px_{pi} \tag{8.1}$$

其中 $x_i = (1, x_{1i}, x_{2i}, \cdots, x_{pi})'$ 表示自变量向量，$Y_i(x_i)$ 表示模糊因变量，A_m，$m = 1,2,\cdots,p$ 为模糊参数。通常，假定模糊参数 A_m 具有三角形隶属函数。因此，模糊因变量 $Y(x_i)$ 的隶属函数也属于三角形形式。

对式(8.1)中的模糊参数 A_m 进行估计时，通常做法是，先假定 A_m 的隶属函数为三角形形式。A_m 的隶属函数可以表达如下

$$\mu_{A_m}(t) = \max\left\{1 - \frac{|t - c_m|}{s_m}, 0\right\}, \quad -\infty < t < \infty \tag{8.2}$$

其中，c_m 表示三角形的中点，s_m 表示三角形的分布半径，而 c_m，s_m 与隶属函数 $\mu_{A_m}(t)$ 的关系，如图 8.1 所示。

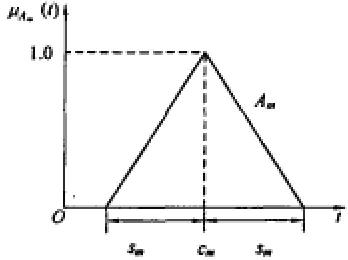


图 8.1 c_m, s_m 在 $\mu_{A_m}(t)$ 中的位置与分布

为了表示方便起见，我们可令模糊参数 $A_m = < c_m, s_m >$，则式(8.2)的模糊数据回归可改写成

$$Y(x_i) = \langle c_0, s_0 \rangle + \langle c_1, s_1 \rangle x_{1i} + \langle c_2, s_2 \rangle x_{2i} + \cdots + \langle c_p, s_p \rangle x_{pi} \quad (8.3)$$

因此，模糊因变量 $Y(x_i)$ 的隶属函数也属于三角形形式，可表示成

$$\mu_{Y(x_i)}(t) = \max\left\{1 - \left| t - \sum_{m=0}^{p} c_m x_{mi} \right| \Big/ \sum_{m=0}^{p} s_m \mid x_{mi} \mid, 0\right\}, \ -\infty < t < \infty \quad (8.4)$$

关于式(8.4)的详细证明，可参看田中、上岛和浅井(Tanaka, Uejima and Asai, 1982)。

在上述模糊数据回归中，模糊参数 A_m 是未知的，这正是需要我们估计的内容。但是，前面已经假定，A_m 的隶属函数为三角形形式。因此，只要能够估计出三角形的中点 c_m 与半径 s_m，就可以获得 A_m 隶属函数的估计值。

8.3 模糊数据回归的参数估计

想要推断模糊数据回归模型，即估计模糊数据回归参数，估计方法有二种。一种是利用线性规划的技巧，来寻求一组能将实数值样本(x_i, y_i)或模糊数样本(x_i, Y_i)全部包含住的区间，再将三角形隶属函数套用在此区间上，如此可求出一组模糊回归估计量，此种方法最常使用。

另一种估计方法则是采用最小二乘法的原则，首先把样本整理成上边界点与下边界点二部分，各自利用上边界点与下边界点求出二条最小二乘法直线，以这两条直线作为三角形隶属函数的边界。对此方法更详细的讨论，可参看吴柏林和曾能芳(1996)。与线性规划估计方法相比，最小二乘法比较符合误差随机分布的精神，此法也是本章讨论的重点。

(1) 线性规划法

设观测到的模糊样本为(x_i, Y_i)，其中 x_i 表示模糊自变量向量，Y_i 表示观测到的样本模糊数，样本模糊数 Y_i 的隶属函数也属于三角形形式，可表达成

$$\mu_{Y_i}(t) = \max\left\{1 - \frac{|t - y_i|}{r_i}, 0\right\}, \quad -\infty < t < \infty \tag{8.5}$$

其中，y_i 表示三角形隶属函数的中点，r_i 则是三角形隶属函数的分布半径，可设 $Y_i = \langle y_i, r_i \rangle$。我们利用 y_i，r_i 对式(8.3) 未知的中点 c_m 与半径 s_m 进行估计。在进行估计之前，首先定义模糊样本 Y_i 与模糊因变数 $Y(x_i)$ 之间的拟合值(Fitted Value) 如下。

【定义 8.1】　拟合值

设(x_i, Y_i)为一组模糊样本，$Y(x_i)$ 表示模糊因变数。若$[Y_i]^h$，$[Y(x_i)]^h$ 分别表示 Y_i，$Y(x_i)$ 的 h 截集(h - cut)，即

$$[Y_i]^h = \{t \mid \mu_{Y_i}(t) \geqslant h\}$$

$$[Y(x_i)]^h = \{t \mid \mu_{Y(x_i)}(t) \geqslant h\}$$

则在第 i 个样本之下，使得$[Y_i]^h \subset [Y(x_i)]^h$ 成立的最大h 值称为第i个样本拟合值，用 h_i 表示。

样本拟合值是用来衡量样本值与估计值之间拟合程度的重要参考指标。若样本拟合值 h_i 愈大，则代表样本 Y_i 和模糊因变数 $Y(x_i)$ 之间吻合得愈好。Y_i，$Y(x_i)$ 以及 h_i 的关系如图 8.2 所示。

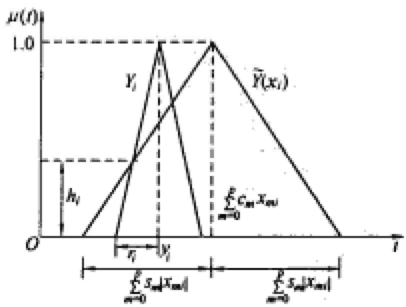


图 8.2　Y_i，$Y(x_i)$ 以及 h_i 的对应图

【性质 8.1】

在定义(8.1) 的条件下，拟合值 h_i 可改写成

$$h_i = 1 - \frac{\left| y_i - \sum_{m=0}^{p} c_m x_{mi} \right|}{\sum_{m=0}^{p} s_m \mid x_{mi} \mid - r_i}, \forall i \in \mathbf{N} \tag{8.6}$$

在估计 c_m, s_m 之时，我们常常会限制所有的 h_i 要大于某个特定值 H，即

$$h_i \geqslant H, \forall i \in \mathbf{N} \tag{8.7}$$

其中 H 由使用者自行设定。因此，可保证样本 Y_i 和模糊因变量 $\tilde{Y}(x_i)$ 之间的吻合程度皆能保持在水平 H 之上。由式(8.7) 可推导出两条约束不等式

$$y_i \geqslant \sum_{m=0}^{p} c_m x_{mi} - (1 - H) \sum_{m=0}^{p} s_m \mid x_{mi} \mid + (1 - H) r_i \tag{8.8}$$

$$y_i \leqslant \sum_{m=0}^{p} c_m x_{mi} + (1 - H) \sum_{m=0}^{p} s_m \mid x_{mi} \mid - (1 - H) r_i \tag{8.9}$$

此外，我们希望估计得到的模糊因变量 $\tilde{Y}(x_i)$ 其三角形隶属函数的半径(图 8.2) 愈小愈好，如此隶属度的信息才能更集中。因此，将对所有 x_i 的三角形隶属函数的半径加总，作为模糊参数估计的目标函数，其定义如下。

【定义 8.2】　模糊数据回归参数目标函数

模糊数据回归参数目标函数(Objective Function) O_f 定义为

$$O_f = \sum_{i=1}^{n} \sum_{m=0}^{p} s_m \mid x_{mi} \mid \tag{8.10}$$

其中 $\sum_{m=0}^{p} s_m \mid x_{mi} \mid$ 表示 $Y(x_i)$ 的隶属函数的半径，$\sum_{i=1}^{n} \sum_{m=0}^{p} s_m \mid x_{mi} \mid$ 则是对所有 x_i 代表 $Y(x_i)$ 的隶属函数半径的总和。

在满足式(8.8)，(8.9) 的条件下，我们希望求出可使 O_f 为极小的 c_m, s_m。一般是采用线性规划方法求解出 c_m, s_m，由此可求出模糊参数 $A_m = \langle c_m, s_m \rangle$ 的估计量。

(2) 最小二乘法

上面，我们曾作出"$h_i \geqslant H, \forall i$" 的限制，这是因为研究者相信：样本 Y_i 会表现出相当的代表性，其与模糊因变量 $Y(x_i)$ 之间的拟合程度至少要大于某个数值(H)，所以在此假设下每组样本 Y_i 和模糊因变数 $Y(x_i)$ 之间的拟合值 h_i 均要大于 H。此外，由于定义了 h_i，使得 Y_i 的 h 截集$[Y_i]^h$ 必定会被 $Y(x_i)$ 的 h 截集$[Y(x_i)]^h$ 所包含，因此，模糊数据回归的估计其实是在求解一组能将全部样本概括住的分布区间，并将此估计区间设计成 $Y(x_i)$ 隶属度的散布范围。

通常，对线性规划法的认识是：具有代表性的样本其 h 截集一定得落在 $Y(x_i)$ 的 h 截集之内。一旦采用这种观点，便表示$[Y_i]^h$ 仅能落在$[Y(x_i)]^h$ 的内

部(因为$[Y_i]^h \subset [Y(x_i)]^h$)。一旦出现离群值(Outlier,又称为奇异值),则会得到一组范围变宽的区间,因而,区间的参考价值作用就不大。为了让估计过程具备稳健性,可以利用最小二乘法,此法的思想是:具有代表性的样本 Y_i 其h 截集$[Y_i]^h$ 应该要尽量接近模糊因变量 $Y(x_i)$ 的 h 截集$[Y(x_i)]^h$,即$[Y_i]^h \approx [Y(x_i)]^h$。因此,$[Y_i]^h$ 不一定要落在$[Y(x_i)]^h$ 之内,其中应允许"$[Y_i]^h \not\subset [Y(x_i)]^h$"的情况发生。

根据最小二乘法的观点,使用 h_i 来估计模糊参数比较冒险,还要重新解释样本 Y_i 所蕴含的意义。前面曾假定,样本模糊数 $Y_i = \langle y_i, r_i \rangle$ 属于三角形式隶属函数,所以样本 Y_i 的隶属度是分布在区间$[y_i - r_i, y_i + r_i]$ 上,其中 $y_{Li} = y_i - r_i$ 表示左端点,$y_{Ri} = y_i + r_i$ 表示右端点。样本 Y_i 的左右端点不太可能与 $Y(x_i)$ 的左右边界完全吻合,这是由于测量误差导致样本和 $Y(x_i)$ 之间有所差错,然而两者之间仍有紧密的关系。我们认为,样本的左右端点 y_{Li}, y_{Ri} 隶属于 $Y(x_i)$ 的程度有一定的水平,假定此水平为 H,因此,假定 y_{Li}, y_{Ri} 隶属于 $Y(x_i)$ 的隶属度会尽量地接近 H 其间的关系如下所示

$$\mu_{Y_{(x_i)}}(y_i - r_i) \approx H \tag{8.11}$$

$$\mu_{Y_{(x_i)}}(y_i + r_i) \approx H \tag{8.12}$$

所谓尽量接近 H 是可以允许隶属度大于H 或小于H,而不像线性规划法那样要求一定大于 H。

根据前面对样本的解释以及式(8.11),(8.12) 的限制条件,利用最小二乘法,可得到一套较符合统计观点的估计程序。用所有样本的左端点$\{(x_i, y_{Li}) \mid i = 1,2,\cdots,n\}$ 拟合一个回归直线方程,称之为左端直线估计方程

$$y_L(x_i) = \hat{L}_0 + \sum_{m=1}^{p} \hat{L}_m x_{mi},\ i = 1,2,\cdots,n \tag{8.13}$$

其中 $\hat{L}_m$ 表示回归系数,再以所有的右端点$\{(x_i, y_{Ri}) \mid i = 1,2,\cdots,n\}$ 拟合右端直线估计方程

$$y_R(x_i) = \hat{R}_0 + \sum_{m=1}^{p} \hat{R}_m x_{mi},\ i = 1,2,\cdots,n \tag{8.14}$$

前面曾假定样本的左右端点隶属于 $\tilde{Y}(x_i)$ 的隶属度会尽量地接近 H,所以,我们以这两条估计量来估计 $\tilde{Y}(x_i)$ 之 H 截集(H - cut) 的左右边界,即

$$[\tilde{Y}(x_i)]^H = \left[\hat{L}_0 + \sum_{m=1}^{p} \hat{L}_m x_{mi}, \hat{R}_0 + \sum_{m=1}^{p} \hat{R}_m x_{mi}\right] \tag{8.15}$$

由此可得,$\tilde{Y}(x_i)$ 的隶属函数为

$$\mu_{\tilde{Y}(x_i)}(t) = \max\left\{1 - \left| t - \sum_{m=0}^{p}\left(\frac{\hat{R}_m + \hat{L}_m}{2}\right)x_{mi}\right| \Big/ \sum_{m=0}^{p}\left(\frac{\hat{R}_m - \hat{L}_m}{2(1-H)}\right)|x_{mi}|, 0\right\} \tag{8.16}$$

其中 $x_{i0} = 1$，而且模糊参数 A_m 的隶属函数为

$$\mu_{A_m}(t) = \max\left\{1 - \frac{\left|t - \left(\frac{\hat{R}_m + \hat{L}_m}{2}\right)\right|}{\left(\frac{\hat{R}_m - \hat{L}_m}{2(1-H)}\right)}, 0\right\} \tag{8.17}$$

即

$$A_m = \left\langle \frac{\hat{R}_m + \hat{L}_m}{2}, \frac{\hat{R}_m - \hat{L}_m}{2(1-H)} \right\rangle \tag{8.18}$$

其中$\frac{\hat{R}_m + \hat{L}_m}{2}$表示三角形底边中点，$\frac{\hat{R}_m - \hat{L}_m}{2(1-H)}$表示分布半径。由此可得，一组估计的模糊数据回归模型

$$\begin{aligned}\tilde{Y}(x_i) = &\left\langle \frac{\hat{R}_0 + \hat{L}_0}{2}, \frac{\hat{R}_0 - \hat{L}_0}{2(1-H)} \right\rangle + \left\langle \frac{\hat{R}_1 + \hat{L}_1}{2}, \frac{\hat{R}_1 - \hat{L}_1}{2(1-H)} \right\rangle x_{1i} + \\ &\left\langle \frac{\hat{R}_2 + \hat{L}_2}{2}, \frac{\hat{R}_2 - \hat{L}_2}{2(1-H)} \right\rangle x_{2i} + \cdots + \\ &\left\langle \frac{\hat{R}_p + \hat{L}_p}{2}, \frac{\hat{R}_p - \hat{L}_p}{2(1-H)} \right\rangle x_{pi}\end{aligned} \tag{8.19}$$

值得注意的是，当样本值为对称形态时，左右回归估计方程其拟合结果是一样的，只有常数项不同。此时我们可以考虑用个别对应项的权值除以常数项的分布半径作为个别对应项的分布半径，即分布半径

$$s_m = \frac{c_m}{s_0}, \quad m = 1, 2, \cdots, p \tag{8.20}$$

而利用最小二乘法找出模糊数据回归估计方程，运算程序比线性规划更简化一些。

在式(8.7)中，线性规划方法是将 H 看做样本与总体的拟合值。但是，最小二乘法所设的 H 是指样本端点 y_{Li}，y_{Ri} 隶属于因变数 $Y(x_i)$ 的程度。显然，其端点隶属于 $Y(x_i)$ 的程度本来不应太高。若 H 设置得愈大，表示样本的端点与因变量的端点差距愈大，样本与总体的隶属函数反而吻合得愈差。其实样本和总体的吻合程度还得由实际情况来判断。一般地说，H 尽可能不要设得过大。因为过大的 H 值将导致 $Y(x_i)$ 的隶属函数的分布范围过于宽广，可以违背自然经验法则。

其实线性规划程序对样本的观念是很保守的，它规定样本必须落在设定的

范围之内。而最小二乘法的解释方法已将样本的设限放宽，这种前提很适合套用在抽样观点比较开放的程序上，然而，此种抽样程序将造就较大的样本离异性，出现离群值的可能性也跟着增大，受到离群值的影响，用最小二乘法求出的估计量将会产生严重的偏倚，以致无法精确估算出模糊参数 $A_m = \langle c_m, s_m \rangle$。也就是说，最小二乘法的估计程序不够稳健，易受离群值的影响，这正是最小二乘法的缺点所在。

8.4　景气对策信号实例

在收集数据时，由于时间的迟延，有时仅能收集到单一时点的数值数据。例如，一个月公布一次的景气指标（Business Index），一季公布一次的国内生产总值（GDP）等等。利用此种方式收集到的数据，其表达出的信息看似真实，实际上它仅表达出某一范围区间的特定位置。它并没有清楚地记录下整个范围的变动状况，比如，景气动向虽是一个月仅记录一次，其实在一个月三十天当中，景气动向每天都在变化。

这类型的数据是先累积整个范围下的变动，再加以处理，最后以单一数值的面貌呈现出来。我们认为，此数据并非具有单一数值的意义，它实际所表达的应是某一范围区间的数值分布状况。也就是说，该数据具有模糊的特性。在这情况下，如果利用单一时间点的数据来建立模型，则无法表现该经济行为的真实面貌，也可能扩大预测结果和实际状态之间的误差，在实证研究过程中应予以慎重考虑。由鉴于此，本节将尝试以模糊回归分析，说明中国台湾景气对策信号和各种金融、经济方面的不同层面指标之间的因果关系，并运用最小二乘法来估计模糊数据回归参数。

为了认识和掌握国家经济建设的趋势、分析国内外经济动向，由中国台湾有关当局每月报到中国台湾的景气指标，其中包括景气对策信号。景气对策信号系由货币供给 M1B，放款金额，票据交换金额，股价指数，制造业新接订单，海关出口值，工业生产指数，制造业成品存货率，非农业部门就业率等九项变动率指标编制而成。综合当期的九项指标，计算其景气综合判断分数后，以数值落入的范围决定其景气对策灯号。共有五种信号灯表示景气状况，红灯表示景气过热，黄红灯表示景气微热，绿灯表示景气稳定，黄蓝灯表示景气欠佳，蓝灯表示景气衰退。

景气对策信号是具有模糊特性的资料，我们选用模糊数据回归来建立此经济模型。九项指标当中，我们挑选货币供给年增长率（X_1），海关出口值年增长

率(X_2),工业生产指数年增长率(X_3)三项指针作为自变量,景气对策灯号则是模糊因变量(Y)。本实例数据是从1990年10月收集至1998年1月,以月为单位,共计88笔。

景气对策信号本来是由九项指标依特定公式所组成,上述三项指针与景气对策信号间的关系似乎已确立,何必再研究这三项指针和景气对策信号间的关系呢?我们建立此回归模型的目的是重新想以模糊的观点,呈现出各项景气指标的另一种面貌。景气对策信号的表示法不光只是红、黄红、绿、黄蓝、蓝五种而已,它应表示成受自变量影响的模糊集合,如此才能清楚呈现景气对策信号的含义。此外,不把九项指针全部列为自变量,要想以更精简的方式来表达指标间的因果关系,这才符合回归模型的精神。

设景气对策信号 $Y(x_i)$ 的模糊数据回归模型为

$$Y(x_i) = A_0 + A_1x_{i1} + A_2x_{i2} + A_3x_{i3}$$

其中,x_{i1} 表示货币供给年增长率,x_{i2} 表示海关出口值年增长率,x_{i3} 表示工业生产指数年增长率(单位皆为%),而 $A_j = \langle c_j, s_j \rangle$ 表示模糊回归系数。对于88笔数据,我们将景气对策信号视为样本模糊数 Y_i,并依据下面灯号与景气分数的对应关系

灯号	景气分数范围
红	38 ~ 45
黄红	32 ~ 37
绿	23 ~ 31
黄蓝	17 ~ 22
蓝	9 ~ 16

将样本灯号化成分数区间,由此可得到各笔数据的左右端点 y_{Li}, y_{Ri}。

我们曾经提过,样本的端点隶属于 $Y(x_i)$ 的程度值 H 应由实际情况来判断。在景气信号的实例当中,样本与总体的吻合程度到底达到多少?即 H 值应该设成多少,才能符合实际状况?根据以往的经验,H 通常介于0.1至0.5间。就本实例而言,我们取一个保守值 $H = 0.3$ 为准;也就是说,希望样本的左右端点隶属于 $Y(x_i)$ 的隶属度会尽量地接近0.3。

因此,根据样本的左右端点 y_{Li}, y_{Ri},可以得到,$Y(x_i)$ 之 H 截集的左右边界估计量。运用最小二乘法,用所有样本的左端点(x_i, y_{Li}) 拟合一回归直线,得到左端直线估计方程

$$y_L(x_i) = \sum_{j=0}^{3} L_j x_{ij} = 14.5 + 0.45x_{i1} + 0.12x_{i2} + 0.18x_{i3}$$

此回归拟合的相关信息如下：

回归参数	估计系数 $\hat{L}_j$	标准偏差 $\hat{\sigma}(\hat{L}_j)$	t 值	p 值
L_0	14.50	0.646	22.45	0.00
L_1	0.45	0.050	8.96	0.00
L_2	0.12	0.032	3.63	0.00
L_3	0.18	0.071	2.56	0.01
				$\hat{\sigma} = 3.06$　$R^2 = 61.0\%$

再以样本的右端点(x_i, y_{Ri})拟合右端直线估计方程，得到

$$y_R(x_i) = \sum_{j=0}^{3} R_j x_{ij} = 20.2 + 0.55x_{i1} + 0.13x_{i2} + 0.21x_{i3}$$

此回归拟合的相关信息如下：

回归参数	估计系数 $\hat{R}_j$	标准偏差 $\hat{\sigma}(\hat{R}_j)$	t 值	p 值
R_0	20.20	0.678	29.85	0.00
R_1	0.55	0.052	10.48	0.00
R_2	0.13	0.034	3.70	0.00
R_3	0.21	0.075	2.83	0.01
				$\hat{\sigma} = 3.21$　$R^2 = 67.0\%$

根据式(8.18)模糊参数A_j的计算公式，可求得

$$A_0 = \langle 17.369, 4.092 \rangle, A_1 = \langle 0.500, 0.073 \rangle$$

$$A_2 = \langle 0.121, 0.006 \rangle, A_3 = \langle 0.196, 0.021 \rangle$$

由此可得，一组估计的模糊数据回归模型

$$\bar{Y}(x_i) = \langle 17.368\,7, 4.091\,9 \rangle + \langle 0.500, 0.073 \rangle x_{i1} + \langle 0.121, 0.006 \rangle x_{i2} + \langle 0.196, 0.021 \rangle x_{i3}$$

我们用此模糊估计方程来表示景气对策信号($Y(x_i)$)和货币供给年增长率(x_1)、海关出口值年增长率(x_2)、工业生产指数年增长率(x_3)之间的因果关系。

一旦有了前面的模糊数据回归估计方程后，未来对景气对策信号的表示方法就不再仅局限于红、黄红、绿、黄蓝、蓝这五种颜色，也就是说，我们看待景气

对策信号时，不再仅有"颜色"的观念而已。举例说明，1997 年 10 月的货币供给年增长率(x_1)为 12.87，海关出口值年增长率(x_2)为 5.9，工业生产指数年增长率(x_3)为 11.05(单位皆为 %)，当时的景气对策信号原用绿灯表示。若将三项自变量指标代入式(8.16)，可得到当期景气对策信号($\tilde{Y}(x_i)$)的隶属函数：$\mu_{\tilde{Y}(x_i)}(T) = \max\{1 - |t - 26.27|/5.28, 0\}$。并将其列示于图 8.3。

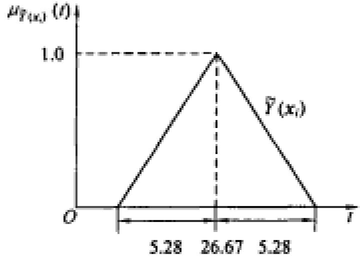


图 8.3 1997 年 10 月景气对策信号的隶属函数

其中 t 表示景气分数，不过它是每个时段皆存在的景气分数，而非有关当局所定义每个月计算一次的景气分数。我们重新用此隶属函数来代表 1997 年 10 月的景气对策信号。这样就可脱离单纯的灯号表示法，此隶属函数更能充分地表示出景气信号所隐含的信息。

8.5 篮球比赛攻防策略实例

篮球比赛的胜负取决于得失分的多寡，且由得失分的情况亦能看出其攻防技术。在一般的比赛记录中，二分球，三分球以及罚球对得分的贡献是很明确的，但是进攻篮板、防守篮板、抄截、失误、犯规等对得失分的影响则无法给出明确的计数。因此，本例子应用模糊统计分析，参考记录员数据与各级教练的意见，分析攻防记录在比赛得失分上适当的模糊权重。然后，应用模糊回归分析找出影响比赛胜负的重要因素，并且用语言变量表示出来，以此提供篮球教练从事训练工作和攻防策略参考。

我们以 1999 年日月光男子甲组篮球联赛作实例分析。数据来源：某个篮协 1999 年 6 月公布的预赛资料。我们针对下面几项作分析和研究。

(1) 通过对 10 位甲组篮球教练的问卷调查数据，求出各攻防项目在得失分上的隶属函数。整理如表 8.1 所示。

表 8.1　各攻防项目对于得失分的隶属度

攻防项目 \ 论域	0 分	1 分	2 分	3 分
进攻篮板球	0.21	0.19	0.39	0.22
防守篮板球	0.21	0.20	0.34	0.25
助攻	0.10	0.23	0.35	0.32
抄截	0.13	0.21	0.40	0.26
阻攻	0.18	0.23	0.36	0.24
攻防项目 \ 论域	0 分	-1 分	-2 分	-3 分
失误	0.14	0.24	0.39	0.23
犯规	0.16	0.23	0.38	0.23

(2) 以隶属函数作为得失分的对应模糊权重，转换出各个攻防项目的模糊分数 k。数据整理，如表 8.2 所示。

表 8.2　攻防项目的模糊分数

攻防项目	模糊分数 k
进攻篮板球 k_4	1.63
防守篮板球 k_5	1.63
助攻 k_6	1.89
抄截 k_7	1.79
阻攻 k_8	1.65
失误 k_9	-1.71
犯规 k_{10}	-1.68

(3) 在收集到的 110 笔数据中，得失分差最多的一场为 57 分，最少的一场为 2 分。我们采用 5 等级分法，将得失分差化为语言变量 L = {不堪一击(1)，略逊一筹(2)，旗鼓相当(3)，略胜一筹(4)，实力超强(5)}，其对应关系如表 8.3。

表 8.3　得失分差和语言变量的对应关系

语言变量 Y_i	不堪一击(1)	略逊一筹(2)	旗鼓相当(3)	略胜一筹(4)	实力超强(5)
得失分差	-57 : -35	-34 : -12	-11 : 11	12 : 34	35 : 57

(4) 我们使用 10 个攻防项目{三分球命中数、二分球命中数、罚球命中数、进攻篮板球、防守篮板球、助攻、抄截、阻攻、失误、犯规} 的对应模糊分数{$k_1, k_2, \cdots, k_{10}$} 乘以该攻防项目在比赛记录中出现的次数{$f_{1i}, f_{2i}, \cdots, f_{10i}$},作为解释变量{$x_1, x_2, \cdots, x_{10}$}。一旦把语言变量 Y_i 作为模糊样本,由模糊数据回归方法,可求出模糊参数 $A_m = \langle c_m, s_m \rangle$。也就是

$$Y(x_i) = A_0 + A_1 x_{1i} + A_2 x_{2i} + \cdots + A_{10} x_{10i}, i = 1,2,\cdots,n$$

其中 $x_i = (1, x_{1i}, x_{2i}, \cdots, x_{10i})'$。

在 8.3 节我们已提过,个别模糊样本 Y_i 的端点隶属于模糊因变量 $\tilde{Y}(x_i)$ 的程度值 H 应由实际状况来判断。而根据以往的经验,H 通常介于 0.1 至 0.5 间。本实证我们取一个中间值 $H = 0.3$;也就是说,希望模糊样本 Y_i 的端点隶属于模糊因变量 $Y(x_i)$ 的隶属度会尽量地接近 0.3。采用最小二乘法,用所有样本的左端点(x_i, y_{Li}) 拟合一个回归直线,得到左端直线估计方程

$$y_L(x_i) = \sum_{j=0}^{10} L_j x_{ji} = -79.8 + 0.420x_{1i} + 0.574x_{2i} + 1.373x_{3i} + 0.069x_{4i} + 0.728x_{5i} + 0.486x_{6i} + 0.916x_{7i} + 0.868x_{8i} + 0.881x_{9i} + 0.10x_{10i}$$

此回归拟合的相关信息说明由表 8.4 给出。

表 8.4 y_L 模型回归参数的相关统计量值

解释变量	回归估计系数 $\hat{L}_j$	标准偏差 $\hat{\sigma}(\hat{L}_j)$	t 值	p 值
三分球 x_{1i}	0.420	0.205 8	2.04	0.044
二分球 x_{2i}	0.574	0.184 1	3.11	0.002
罚球 x_{3i}	1.373	0.308 7	4.45	0.000
进攻篮板球 x_{4i}	0.069	0.194 5	0.35	0.724
防守篮板球 x_{5i}	0.728	0.170 4	4.27	0.000
助攻 x_{6i}	0.486	0.197 3	2.46	0.016
抄截 x_{7i}	0.916	0.233 1	3.93	0.000
阻攻 x_{8i}	0.868	0.416 4	2.08	0.04
失误 x_{9i}	0.881	0.179 4	4.91	0.000
犯规 x_{10i}	0.100	0.205 5	0.49	0.627
			$\hat{\sigma} = 12.83$	$R^2 = 76.6\%$

再用样本的右端点(x_i, y_{Ri}) 拟合右端直线估计方程,得到

$$y_R(x_i) = \sum_{j=0}^{10} R_j x_{ji} = -57.8 + 0.420x_{1i} + 0.574x_{2i} + 1.373x_{3i} + 0.069x_{4i} +$$

$0.728x_{5i}+0.486x_{6i}+0.916x_{7i}+0.868x_{8i}+0.881x_{9i}+0.10x_{10i}$

此回归拟合的相关信息说明由表 8.5 给出。

表 8.5　y_R 模型回归参数相关统计量值

解释变量	回归估计系数 $\hat{R}_j$	标准偏差 $\hat{\sigma}(\hat{R}_j)$	t 值	p 值
三分球 x_{1i}	0.42	0.206	2.04	0.04
二分球 x_{2i}	0.57	0.184	3.11	0.0 +
罚球 x_{3i}	1.37	0.309	4.45	0.0 +
进攻篮板球 x_{4i}	0.07	0.195	0.35	0.72
防守篮板球 x_{5i}	0.73	0.170	4.27	0.0 +
助攻 x_{6i}	0.49	0.197	2.46	0.02
抄截 x_{7i}	0.92	0.233	3.93	0.00
阻攻 x_{8i}	0.87	0.416	2.08	0.04
失误 x_{9i}	0.88	0.180	4.91	0.00
犯规 x_{10i}	0.10	0.206	0.49	0.63
			$\hat{\sigma}=12.83$	$R^2=76.6\%$

根据式(8.18)与(8.20)的计算公式，可求得模糊参数 A_j 如下

$$A_0=\langle -68.8,15.7\rangle,A_1=\langle 0.42,0.03\rangle,A_2=\langle 0.57,0.04\rangle$$

$$A_3=\langle 1.37,0.09\rangle,A_4=\langle 0.07,0.0+\rangle,A_5=\langle 0.73,0.05\rangle$$

$$A_6=\langle 0.49,0.03\rangle,A_7=\langle 0.92,0.06\rangle,A_8=\langle 0.87,0.06\rangle$$

$$A_9=\langle 0.88,0.06\rangle,A_{10}=\langle 0.10,0.0+\rangle$$

由此得到，一组估计的模糊数据回归模型

$$\begin{aligned}Y(x_i)=&\langle -68.8,15.71\rangle+\langle 0.42,0.03\rangle x_{1i}+\langle 0.57,0.04\rangle x_{2i}+\\&\langle 1.37,0.09\rangle x_{3i}+\langle 0.07,0.0+\rangle x_{4i}+\langle 0.73,0.05\rangle x_{5i}+\\&\langle 0.49,0.03\rangle x_{6i}+\langle 0.92,0.06\rangle x_{7i}+\langle 0.87,0.06\rangle x_{8i}+\\&\langle 0.88,0.06\rangle x_{9i}+\langle 0.10,0.0+\rangle x_{10i}\end{aligned}$$

我们用此模糊估计方程来表示语言变量 $Y(x_i)$ 和各个攻防项目之间的因果关系，并且得到攻防项目的重要性依序为罚球、抄截、失误、阻攻、防守篮板球、二分球命中数、助攻、三分球命中数、犯规及进攻篮板球。

值得注意的是，当对回归系数进行估计时，若影响变量彼此有高相关的现象，就会产生多重共线性的情况，从而造成回归系数的偏倚，并使得回归系数不能反映出真正的预测力。要排除此种疑问，比较常用的方法是作多重共线性诊断。对多重共线性进行检验，一般可考虑两种途径：1. 经验判断；2. 数据分析。有

关篮球攻防技术的探讨文献中，对于本实例所筛选的 10 项影响变量，已作过多次的多重共线性诊断，其结论都是没有共线性的。因此，在本实例中，这些影响变量的选取应是相当合理且有效的过程。

(5) 将$\{x_{1i},x_{2i},\cdots,x_{10i}\}$依照进攻 $O_i=$ {三分球，二分球，助攻，进攻篮板球}；防守 $D_i=$ {防守篮板球，阻攻，抄截}；EQ& 战术 $E_i=$ {罚球，犯规，失误} 分类，以{进攻，防守，战术 &EQ} 作为解释变量，语言变量 Y_i 作为模糊样本，求出模糊参数

$$A_m=\langle c_m,s_m\rangle$$

$$Y(x_i)=A_0+A_1O_i+A_2D_i+A_3E_i,i=1,2,\cdots,n$$

$$O_i=x_{1i}+x_{2i}+x_{4i}+x_{6i}$$

$$D_i=x_{5i}+x_{7i}+x_{8i}$$

$$E_i=x_{3i}+x_{9i}+x_{10i}$$

其中，$x_i=(1,O_i,D_i,E_i)'$，EQ 表示情商。

用样本的左端点(x_i,y_{Li})拟合左端直线估计方程，得到

$$y_L(x_i)=-62.3+0.426O_i+0.827D_i+0.686E_i$$

此回归拟合的相关信息说明由表 8.6 给出。

表 8.6 y_L 模型回归参数相关的统计值

解释变量	回归估计系数	标准偏差 $\hat{\sigma}$	t 值	p 值
进攻面 O_i	0.43	0.07	6.4	0.00
防守面 D_i	0.83	0.13	6.4	0.00
EQ& 战术 E_i	0.69	0.14	4.9	0.00
			$\hat{\sigma}=13.83$	$R^2=71.0\%$

再用样本的右端点(x_i,y_{Ri})拟合右端直线估计方程，得到

$$y_R(x_i)=-40.3+0.426O_i+0.827D_i+0.686E_i$$

此回归拟合的相关信息说明由表 8.7 给出。

表 8.7 y_R 模型回归参数相关的统计值

解释变量	回归估计系数	标准偏差 $\hat{\sigma}$	t 值	p 值
进攻面 O_i	0.43	0.07	6.4	0.00
防守面 D_i	0.83	0.13	6.4	0.00
EQ& 战术 E_i	0.69	0.14	4.9	0.00
			$\hat{\sigma}=13.83$	$R^2=71.0\%$

根据式(8.18)与(8.20)的计算公式，可求得模糊参数 A_j 如下

$$A_0 = \langle -51.3, 15.7\rangle,\quad A_1 = \langle 0.43, 0.03\rangle$$

$$A_2 = \langle 0.83, 0.05\rangle,\quad A_3 = \langle 0.7, 0.04\rangle$$

由此得到，一组估计的模糊数据回归模型

$$Y(x_i) = \langle -51.3, 15.71\rangle + \langle 0.47, 0.03\rangle O_i + \langle 0.83, 0.05\rangle D_i + \langle 0.69, 0.04\rangle E_i$$

根据模糊估计方程，可以得到整合式的排序为：防守对胜负的影响最大；其次为 EQ& 战术；进攻的影响力最小。值得探讨的是，何种攻防项目该分到进攻类，何种攻防项目该分到防守类，何种攻防项目该分到 EQ& 战术类，还需要更进一步研究。

8.6　结　　论

本章研究提出有别于传统的、通过线性规划估计回归参数的方法。因为线性规划是一种比较注重数值运算的估计程序，其观点与统计学的精神有相当的差异。因此，我们利用模糊统计分析与算法，重新解释样本 Y_i 所蕴含的意义。如此样本的解释将更符合实际状况。本章运用最小二乘法，求得模糊因变量的左右边界，进而导出模糊数据回归的估计量，应是一套比较符合统计原则的估计程序。

与一般线性规划的估计技巧相比，本章提供的最小二乘法亦能充分吻合样本的特性，而且运算程序也较为简易。虽然最小二乘法易受离群值的影响，有时可能会产生较大的拟合误差。因而，我们也考虑引用非参数回归估计技巧，以降低离群值的影响力，增加模糊数据参数估计的稳健性。因此，与传统方法相比，本章提出的估计方法更能准确地描述出 X 与 Y 之间的模糊因果关系。

在景气对策信号的实例分析当中，最后导出模糊数据回归估计方程，可以看出，景气对策信号和三项自变量在结构上的相关程度。由此，我们也能表达出景气对策信号所隐含的模糊意识。在此实例中，容易看出，与传统回归相比，模糊数据回归具有更切贴现实的解释能力。另外，以模糊的观念来解释景气信号的数据结构，过去某些真实且重要的、但又经常被人们忽略的信息就可清楚地呈现出来。

另一方面，在传统的回归分析中，最小二乘法所求出的估计量为最佳线性无偏估计量(BLUE，Best Linear Unbiased Estimate)。但是，本章研究所引用最小二乘法求得的模糊数据回归估计量是否也具有类似的性质呢？在模糊分析框架下，又该如何重新定义像 BLUE 这类的估计特性呢？在本章，我们虽然提出一套

完善的模糊数据回归参数估计程序，但是对此估计程序的统计性质并未详细推导其数理特性。因此，对此方向的进一步研究发展将是未来研究的方向及重点。

而在篮球比赛攻防策略模糊统计分析中，参考各级教练的意见，统计攻防记录在比赛得失分上适当的模糊权重。我们考虑如何将不同教练的模糊印象与看法，转化为排序权重以得到各攻防项目对应的模糊分数。然后，使用记录员数据与模糊回归分析，找出影响比赛胜负的重要因素，并且用语言变量表示出来。因而，自变量应以各项目实际的得失分为计算单位，才能反映出它们对胜负的真正影响。至于进攻篮板球、防守篮板球、助攻、抄截、阻攻、失误及犯规，则可以通过模糊权重的设定，将其转换为适当的得失分。这样的软统计分析应该是比较可行且合理的。

第9章　模糊数据样本排序及非参数检验方法

在传统的统计推断方法中，对数字的大小进行排序是一个很重要的步骤。通过排序行为，往往能让我们了解诸如集中趋势、分散情况以及偏态等数据分布的形式。然而，模糊数及模糊区间数的排序却不像明确数字那样容易实施，如何找出简易有效的排序方法，一直是模糊数学界相关学者努力的方向。

针对模糊数据，本章尝试以软计算方法，结合模糊理论，定义出模糊数及模糊区间的排序方法，并利用此种方法找出模糊总体中位数的估计方法。同时，针对传统非参数检验方法无法解决参数假设为模糊数或模糊区间值的缺点，我们提出模糊数据符号检验，模糊数据威尔克森(Wilcoxon)符号秩和检验方法以及模糊数据克鲁斯考尔和沃利斯(Kruskal - Wallisru)检验。由例子及实证数据分析结果显示，我们提出的检验方法能有效解决模糊样本问题。

9.1　模糊数据样本排序

在传统的统计分析方法中，排序是一个很重要的步骤，然而，当样本数据值为模糊数或模糊区间值时，排序的问题便变得复杂且难以解决。关于模糊数的排序，相关的文献很多，弗里林(Freeling,1980)系统地将现有的模糊排序方法依照比较的技术分为直接比较法、语意法等五类。Chen与Hwang(1992)则将模糊排序方法依喜好关系、模糊平均数及分散度、模糊评分、语意表达分成四种。考夫曼和格普塔(Kaufmann and Gupta,1988)针对具三角形隶属函数的模糊数，利用其三顶点的值，提出一个排序三法则；Liou与Wang(1992)则是针对具有凸隶属函数的模糊数，利用积分的方法，比较模糊数的期望值大小来确定模糊数的排序；Cheng(1988)则是利用计算距离的方法来比较模糊数的大小。

对于模糊数的排序方法，有的只适用于具有特殊形式隶属函数的模糊数，有的则是计算起来比较繁琐。因此，本章针对日常生活常见的模糊数据形式，即多值逻辑模糊数及模糊区间数的排序，提出一种有效且计算便捷的排序方法。

传统的统计检验必须对数据分布的形式给予明确的假设，例如，总体分布为正态或近似正态。例如，当我们想要检验两个总体平均数是否有差异时，零假

设(Null Hypotheais,又称为原假设)是“两个平均数相等”,而且通常假定总体服从正态分布。然而,有时我们想要知道的只是两个平均值是否非常逼近,此时传统的检验方法并不适用于这种包含不确定性的假设检验。再比如,针对某股市分析师宣称某股票股价将在23元左右盘整,23元这个数字看似精确,事实上却带有不确定性及模糊性,如果我们打算针对此假设进行检验,那就应视为是一个有关模糊数的假设检验,而传统的统计检验方法对于这样假设并无有效的解决方法。此外,针对某经济研究机构宣称,国内失业率将达到大约3% ~ 5%这样带有不确定性且为区间值的模糊零假设,传统的假设检验也缺乏比较有效可行的检验方法。

由上述可知,模糊统计推断日渐受到重视,这应是复杂的社会现象无法利用传统数值模型充分合理解释的一种自然发展结果。有鉴于此,本章尝试以软计算方法,结合模糊理论,针对多值逻辑模糊数及模糊区间数,提出有效可行的排序方法,并把这种方法应用于符号检验及威尔克森符号秩和检验等非参数检验方法。同时,将此方法应用于实证数据,希望能对相关的研究提供切合实际的有效分析方法。

由于连续型的模糊区间样本,其大小排序的情况远比离散型模糊样本来得更复杂,相关文献也较少提出排序方法。其中,多半将模糊区间数视为一个具有凸隶属函数的模糊数,再以积分的方法找出各模糊数的模糊中心,然后用模糊中心的大小决定各模糊数的排序。如此的排序方法会遭遇几个问题:其一是凸隶属函数的决定会较主观且难以决定,其二是对大样本而言,用积分方法来决定模糊数中心比较繁琐,且实际应用上较为不便。本章我们以均匀分布为基础,结合模糊关系隶属函数,提出下述方法来决定模糊区间样本的大小关系。

【定义9.1】 模糊数据样本排序

设 U 表示一个论域,令 x_{1f}, x_{2f} 为模糊数据样本 x_1, x_2 的反模糊化值,也就是 $x_f = \sum_{i=1}^{k} m_i L_i$ 对离散模糊数据样本 $x = \sum_{i=1}^{k} \frac{m_i}{L_i}$, $x_f = \frac{x_l + x_u}{2}$,对连续模糊数据样本 $x = [x_l, x_u]$,则:

如果 $x_{1f} < x_{2f}$,称 x_1 模糊小于 x_2,记为 $x_1 \prec x_2$;

如果 $x_{1f} = x_{2f}$,称 x_1 模糊等于 x_2,记为 $x_1 \approx x_2$;

如果 $x_{1f} > x_{2f}$,称 x_1 模糊小于 x_2,记为 $x_1 \succ x_2$。

【例9.1】 离散型模糊数据样本排序应用于餐饮业服务满意度调查。

针对哈尔滨市某大学师生对学校餐厅服务是否满意进行调查,将满意程度分为:L_1 = 很不满意,L_2 = 不满意,L_3 = 普通,L_4 = 满意,L_5 = 很满意五种。随

机抽取 6 位学校师生，以模糊问卷方式进行调查，对学校餐厅满意程度的隶属度加以计算统计，如表 9.1 所示。

表 9.1　6 位受访者对应于五等级隶属度的选择

	L_1	L_2	L_3	L_4	L_5	x_f
1	0	0.5	0.5	0	0	2.5
2	1	0	0	0	0	1
3	0	0.7	0.3	0	0	2.3
4	0	0	0.8	0.2	0	3.2
5	0	0	0	0.1	0.9	4.9
6	0	0	0.6	0.4	0	3.4

则 6 位受访者对学校餐厅服务满意程度排序为

$$x_2 < x_3 < x_1 < x_4 < x_6 < x_5$$

【例 9.2】　2006 年毕业生求职潮中，我们调查出北京市某大学 8 位硕士毕业生对年薪水期望的一组模糊区间样本如表 9.2 所示，而去年调查结果的某大学硕士毕业生年薪水期望均数为 3 ~ 4 万元。

表 9.2　8 位硕士毕业生对期望薪水区间值调查

i	x_l	x_u	$x_f = \frac{x_l + x_u}{2}$
x_1	2	3	2.5
x_2	3	4	3.5
x_3	2	5	3.5
x_4	2.6	4	3.3
x_5	3.5	5.5	4.5
x_6	3.6	4.2	3.9
x_7	3.2	5	4.1
x_8	2.6	4.6	3.6

则 8 位硕士生对期望年薪水排序为

$$x_1 < x_4 < x_2 \approx x_3 < x_8 < x_6 < x_7 < x_5$$

9.2 模糊数据中位数的符号检验及应用

符号检验可以说是起源最早的统计检验方法，通常用来检验两个总体 X，Y 是否相等，这里相等的意思是指中位数没有差异的情况。

符号检验也可以推广用来检验单一总体的中位数，或一个时间数列过程的趋向性。例如，若要检验总体的中位数是否为某特定值 M_0 时，其统计假设 H_0: $M_X = M_0$，H_1: $M_X \neq M_0$。我们可由总体随机抽取 n 个样本，然后比较样本值与总体中位数 M_0 的大小。若 $x_i > M_0$，则记以"+"，若 $x_i < M_0$，则记以"-"(舍去 $x_i = M_0$ 的样本)。如果该总体 X 的中位数确实是 M_0，那么出现正号的总数与出现负号的总数应相差不多。若出现"+"的总数与出现"-"的总数相差颇大时，则我们可以说总体 X 的中位数不等于某特定值 M_0。

模糊数据样本的符号检验法，即是考虑将模糊数据样本反模糊化后，再将所得的值与总体中位数 M_0 的大小作比较所得的统计结果。另外，为了简洁叙述检验过程，避免重复说明起见，本章各种检验法介绍大都只叙述双侧检验过程，至于单侧检验的情况均可比照类推。下面，依据模糊数据样本为离散型及连续型，分述其检验步骤如下。

模糊数据样本符号检验程序

(1) 样本：一组随机的成对模糊离散数据样本 (x_i, y_i)，$i = 1, \cdots, n$ 或一组中位数已知的随机模糊离散数据样本。

(2) 统计假设：X，Y 两个总体分布的模糊中位数没有差异，也就是

$$H_0: Fmedian(X) = Fmedian(Y) \quad 对 \quad H_1: Fmedian(X) \neq Fmedian(Y)$$

(3) 统计量：T = 出现 $x_i > y_i$ 的总数。

4. 决策法则：在显著水平 α 下的双侧检验，若

$$P(T \leqslant C_\alpha) = \sum_{i=0}^{C_\alpha} \binom{n}{i} (0.5)^i (0.5)^{n-i} < \alpha/2$$

或若

$$P(T \geqslant C_\alpha) = \sum_{i=0}^{n} \binom{n}{i} (0.5)^i (0.5)^{n-i} < \alpha/2$$

则拒绝 H_0。

当小样本($n \leqslant 25$)时，我们可利用附录 2，查到相对应 α 的临界值。若大样本时，则可应用中心极限定理，由逼近正态分布法则，以求得 α 的临界值。因为在 H_0 下，符号检验统计量 T 具有期望值 $\mu_X = np = n/2$，方差 $\sigma^2 = npq = n/4$

的二项式分布。因此，可使用 Z 检验量

$$Z = \frac{T - \mu_X}{\sigma_X} = \frac{T - (n/2)}{\sqrt{n/4}} = \frac{2T - n}{\sqrt{n}} \tag{9.1}$$

来检验。

【例 9.3】　协和基金会为了探讨夫妻在家庭重要决策中，是否如传言妻子具有较大影响力？于是，随机抽取 17 对夫妻针对家庭中重要的决策情况作统计。

在问卷中要求他们对决策的影响力作一模糊回答，$\left\{x_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}\right\}$，其中影响力分为五个语意变量：1 = 很没影响，2 = 较无影响，3 = 普通，4 = 较有影响，5 = 很有影响等五种。将模糊数据样本反模糊化值后，其统计结果如表 9.3 所示。

表 9.3　夫妻在家庭重要决策中的影响力

样本	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
妻 x_{ij}	3.5	3	4.2	4	2.5	1.5	3.5	2.4	1.5	2.5	3	2.5	2.5	3.5	3.5	3.5	2
夫 y_{ij}	2	2.3	2.2	3	2.5	4.4	2.1	2.4	4.8	3.6	2.2	2	4	2.5	2.6	1	4.5
符号	+	+	+	+	0	−	+	0	−	+	+	+	−	+	+	+	−

H_0: $Fmedian(X) = Fmedian(Y)$，即夫妻在家庭决策中的影响力一样；

H_1: $Fmedian(X) > Fmedian(Y)$，即妻子在家庭决策中的影响力较大。

由表 9.3 发现，正号总数 $T = 11$，$n = 17 - 2 = 15$，相持不记。用右侧检验，在显著水平 $\alpha = 0.1$ 下，查附录 2 得

$$P(T \geq 11) = 1 - B(10;15,0.5) = 1 - \sum_{i=0}^{10} C_i^{15}(0.5)^i(0.5)^{15-i} =$$
$$1 - 0.94 = 0.06 < \alpha = 0.1$$

故而拒绝 H_0，即在显著水平 $\alpha = 0.1$ 下，妻子在家庭重要决策中的影响力较大。

【例 9.4】　连续型模糊数据最大隶属度估计量应用于等候时间调查。

对快餐店业者来说，缩短顾客等候时间是非常重要的，不但可提升顾客对服务的满意度，也可以增加销售金额。某快餐店业者采用开放式问卷，随机抽取 10 位顾客，调查他们可忍受的最长等候时间区段，结果如表 9.4 所示，针对去年委托民调中心调查的结果为可忍受等候时间为 4 ~ 8 min，也就是 $F_{x_0} = [4,8]$。打算检验近来顾客可忍受等候时间与去年是否有差异。顾客可忍受等候时间对各模糊关系的隶属度，如表 9.4 所示。

表 9.4 10 位顾客对等候时间的可忍受时间区段

时间 /min	1	2	3	4	5	6	7	8	9	10
x_l	4	7	2	2	1	1	5	3	4	6
x_u	7	10	5	8	4	2	10	7	9	8
c_i	5.5	8.5	2.5	5	2.5	1.5	7.5	4.5	6.5	7
符号	-	+	-	-	-	-	+	-	+	+

H_0: $Fmedian(X) = Fmedian(Y)$，即顾客可忍受等候时间区间中位数与去年差不多；

H_1: $Fmedian(X) < Fmedian(Y)$，即顾客可忍受等候时间区间中位数与去年有显著差异。

由表 9.4 发现，正号总数 $T = 4, n = 10$。用左右两侧检验，在 $\alpha = 0.1$ 显著水平下，查附录 2 得

$$P(T \leqslant 4) = B(4;10,0.5) = \sum_{i=0}^{4} C_i^{10}(0.5)^i(0.5)^{10-i} = 0.37 > \alpha = 0.1$$

故而接受 H_0，即在 $\alpha = 0.1$ 显著水平下，顾客可忍受等候时间区间中位数与去年差不多。

9.3 模糊数据样本排序方法应用于威尔克森符号秩检验

符号检验法仅仅利用到成对样本之间的差异方向（正或负），并没有考虑其差量的大小。因此，检验过程中，可能会失去一些重要的数据信息。威尔克森（Wilcoxon，1945）提出符号秩检验方法，就是将成对样本间差量大小与符号方向同时纳入考虑之中，也就是对差量越大的配对，给予越大的权值，如此也就提高了检验方法的检验力。

威尔克森符号秩检验方法的概念为：假定我们由随机成对样本 $\{(X_i, Y_i) \mid i = 1, \cdots, n\}$ 中，要检验两个总体 X, Y 是否有差异。令 $d_i = Y_i - X_i$ 为成对 (X_i, Y_i) 的差值，$i = 1, \cdots, n$。对 d_i 取绝对值，再以秩 1 给予 $|d_i|$ 最小值，秩 2 给予 $|d_i|$ 次小值，…，依此类推，以秩 n 给予 $|d_i|$ 最大值。若有数个 $|d_i|$ 值相等，则取其在不同值下所对应秩的平均值。由此可知，全部 $|d_i|$ 秩和为 $n(n+1)/2$。

设 T^+ 为所有 d_i 为正数的秩和，T^- 为所有 d_i 为负数的秩和。由此可知，T^+ 或 T^- 的最小值均可能为 0，且 T^+ 或 T^- 的最大值均可能为 $n(n+1)/2$。因此，在统计假设 $H_0: M_X = M_Y$ 下，若 T^+ 或 T^- 接近 $n(n+1)/4$，则我们可接受 H_0：X, Y 两个总体没有差异。反之，若 T^+ 或 T^- 接近 0 或 $n(n+1)/2$，则我们拒绝 H_0。

模糊数据样本的符号检验法，也就是考虑把模糊样本反模糊化后，再将所得的值与总体中位数 M_0 的大小作比较所得的统计结果。

模糊数据样本威尔克森符号秩检验程序

(1) 样本：一组随机的成对模糊数据样本 (x_i, y_i)，$i = 1, \cdots, n$，或一组中位数已知的随机模糊样本。(x_{if}, y_{if}) 为其相对的反模糊化值。$d_i = y_{if} - x_{if}$，$i = 1, \cdots, n$。

(2) 统计假设：X, Y 两个总体分布的模糊中位数没有差异，即

$$H_0: Fmedian(X) = Fmedian(Y) \quad 对 \quad H_1: Fmedian(X) \neq Fmedian(Y)$$

(3) 统计量：$T = T^+$ 与 T^- 中较小的值。

(4) 决策：在显著水平 α 下的左侧检验，查附录 7，若 $T < T_{\alpha/2}$（临界值），则拒绝 H_0。

当小样本（$n \leqslant 25$）时，我们可利用附录 2 查到相对应 α 的临界区。当大样本时，可以证明，威尔克森符号秩检验量 T 的分布会趋近于正态分布，其期望值与方差为

$$E(T) = \frac{n(n+1)}{4}, \quad \sigma_T^2 = \frac{n(n+1)(2n+1)}{24} \tag{9.2}$$

因此，可以使用统计量 $Z = \dfrac{T - E(T)}{\sigma_T}$ 来检验。

【例 9.5】　北京市某大学心理系幼教心理学专家想要探讨双胞胎中，首先出生者是否比后出生者更有进取心。于是，对选取的 12 对双胞胎给予心理测验，并以模糊隶属度记录其模糊分数 (x_i, y_i)。然后，将其反模糊化得到 (x_{if}, y_{if})。结果如表 9.5 所示。较高分者表示比较有进取心。

表 9.5　双胞胎中首先出生者与后出生者心理测验结果

双胞胎	1	2	3	4	5	6	7	8	9	10	11	12
先出生者	86	71	77	68	91	72	77	91	70	71	88	87
后出生者	88	77	76	64	96	72	65	90	65	80	81	72
d_i	2	6	-1	-4	5	0	-12	-1	-5	9	-7	-15
$\lvert d_i \rvert$ 的等级	3	7	1.5	4	5.5	-	10	1.5	5.5	9	8	11

利用威尔克森符号秩检验法检验，在显著水平 $\alpha = 0.05$ 下，是否双胞胎中首先出生者比后出生者更有进取心。

统计假设 H_0：双胞胎中首先出生者不比后出生者更有进取心；

对 H_1：双胞胎中首先出生者比后出生者更有进取心。

由表 9.5 可得，正号总和 $T^+ = 3 + 7 + 5.5 + 9 = 24.5$，负号总和 $T^- = 1.5 + 4 + 10 + 1.5 + 5.5 + 8 + 11 = 41.5$。用单侧检验，在统计假设 H_0 下，查附录 7 可得，当 $n = 11$，$\alpha = 0.05$ 时，$T_{0.05} = 14 < T^+ = 24.5$。因此，我们接受 H_0，即双胞胎中首先出生者不比后出生者更有进取心。

9.4 模糊数据样本排序方法应用于威尔克森秩和检验

威尔克森秩和检验仍可用来检验两组独立样本，其所来自总体的中位数是否相等。在非参数统计检验法中，一般地说，威尔克森秩和检验是应用最广的一种检验法。威尔克森秩和检验与一般所谓的曼－惠特尼(Mann-Whitney)检验法原理是一致的(可参看孙山泽《非参数统计讲义》第 11～12 页，或者王静龙和梁小筠《非参数统计分析》第 111～112 页)。

假定我们从两个独立总体 X，Y 随机抽出 m，n 个模糊样本，然后将此样本反模糊化，最后依照样本值大小，将所有 $m + n = N$ 个混合样本由小而大排序。也就是，在混合样本中，最小的混合值给予 1，次小者给予 2，…，依此类推，最大的混合值给予 N。若有混合值相等者，则取其在不同值下所对应秩的平均值。设 $R(X_i)$，$R(Y_j)$ 分别表示 X_i，Y_j 在混合样本中所对应的秩，则两组样本秩的总和分别为

$$W_X = \sum_{i=1}^{m} R(X_i) \quad 与 \quad W_Y = \sum_{i=1}^{n} R(Y_i)$$

可知 W_X 的最小可能值为 $1 + \cdots + m$，最大可能值为 $(n + 1) + \cdots + (n + m)$。$W_Y$ 的最小可能值为 $1 + \cdots + n$，最大可能值为 $(m + 1) + \cdots + (n + m)$。

因此，如果总体 X 的样本值比总体 Y 的样本值来得大，即总体 X 的样本值集中于秩高的区域，那么 W_X 值将比较大。如果总体 X 的样本值比总体 Y 的样本值来得小，即总体 X 的样本值集中于秩低的区域，那么 W_X 值将比较小。在这两种情况下，则我们会拒绝统计假设 $H_0: M_X = M_Y$。

举一个实例来说，我们有两组样本分别来自初中联考班 X 与自学班 Y 的学生，其反模糊化测验成绩如下。联考班：70，75，85，90，自学班：55，60，80。我们将

混合样本依大小排秩如下：

观测值	55	60	70	75	80	85	90
总体	Y	Y	X	X	Y	X	X
秩	1	2	3	4	5	6	7

则
$$W_X = 3+4+6+7 = 20, W_Y = 1+2+5 = 8$$

威尔克森秩和检验程序

(1) 数据：两组独立总体 X,Y 的随机模糊数据样本 $x_1,\cdots,x_m,y_1,\cdots,y_n$，$m+n=N$。$x_{1f},\cdots,x_{mf},y_{1f},\cdots,y_{nf}$ 为其相对的反模糊化值。

(2) 统计假设：X,Y 两总体分布的中位数没有差异，也就是

$$H_0: M_X = M_Y \quad 对 \quad H_1: M_X \neq M_Y$$

(3) 统计量：$T = W_X = \sum_{i=1}^{m} R(X_i)$。

(4) 决策：在显著水平 α 下的双侧检验，查附录 8，若 $W_L \leq T \leq W_U$，则接受 H_0。

当 $m,n(m>10,n>10)$ 很大时，可以证明，威尔克森秩和检验统计量 T 的分布会趋近于正态分布，其期望值与方差为

$$E(T) = E\left(\sum_{i=1}^{m} R(X_i)\right) = \frac{m(N+1)}{2} \tag{9.3}$$

$$\sigma_T^2 = Var\left(\sum_{i=1}^{m} R(X_i)\right) = \frac{mn(N+1)}{12} \tag{9.4}$$

因此，可使用统计量 $Z = \dfrac{T-\mu_T}{\sigma_T}$ 来检验。

【例 9.6】　黑龙江省某调查公司想要了解游客对本省旅游胜地：镜泊湖与五大连池风景区的满意度情况。于是，在 2006 年 7 月的某假日随机各抽取数名游客，就其景观、设施、环保、服务、交通等项作隶属度评分，得到反模糊化值分值，如表 9.6 所示。

表 9.6　镜泊湖与五大连池风景区的满意度

镜泊湖 x_f	40.2	38.4	32	45.3	33.4	45.8		
五大连池风景区 y_f	28.7	35.9	32	42.2	47.7	24.3	33.6	41.4

利用威尔克森秩和检验，在显著水平 $\alpha=0.05$ 下，统计假设为：

H_0：游客对镜泊湖与五大连池风景区满意度相同，即 $M_X = M_Y$。

H_1:游客对镜泊湖与五大连池风景区满意度不同,即 $M_X \neq M_Y$。

将游客对镜泊湖与五大连池风景区的满意度混合排序如下:

满意度	24.3	28.7	32	32	33.4	33.6	35.9	38.4	40.2	41.4	42	45.3	45.8	47.7
总体	Y	Y	Y	X	X	Y	Y	X	X	Y	Y	X	X	Y
秩	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14

由上表可得,$W_X = 3.5 + 5 + 8 + 9 + 12 + 13 = 50.5$。在统计假设 H_0 下,查附录8可得,当 $m = 6, n = 8, \alpha = 0.05$ 时,$(W_L, W_U) = (29,61)$ 此区间包含50.5。因此,我们接受 H_0,即游客对镜泊湖与五大连池风景区满意度相同。

9.5 模糊数据样本排序方法应用于克鲁斯考尔和沃利斯秩检验

克鲁斯考尔和沃利斯(Kruskal-Wallis)秩检验法是一种最常用于检验多组独立总体间是否有差异的方法。假定我们有 k 组独立的总体,而每个总体的模糊观察值有 n_i 个,$i = 1,\cdots,k$。现在,依照反模糊化值的大小将所有 $n_1 + n_2 + n_3 + \cdots + n_k = N$ 个反模糊化值排序,最小的反模糊化值给予1,次小者给予2,依此类推,最大的反模糊化值观测值给予 N;若有数个反模糊化值相等,则取其在不同值下所对应秩的平均值。

然后,我们对 k 组总体求秩和,以 R_i 表第 i 个模糊总体在混合样本中所对应的秩和,若我们所假设的是 H_0:各模糊总体间无差异,显然,若各组独立模糊总体的样本大小一致时,则所有 R_i 值($i = 1,\cdots,k$)不会过大或过小,但是如果有某个总体的秩和太大(或太小),那么我们拒绝 H_0。

模糊数据克鲁斯考尔和沃利斯秩检验程序

(1) 数据:k 组相互独立模糊数据总体,每个总体有 n_i 个模糊观测值,$i = 1,\cdots,k$,$n_1 + n_2 + n_3 + \cdots + n_k = N$,$x_{ij}$ 为其相对的反模糊化值。

(2) 假设:H_0: $M_1 = M_2 = \cdots = M_k$。

(3) 统计量

$$T = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{1}{n_i}\left[R_i - \frac{n_i(N+1)}{2}\right]^2 = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1)$$

其中 R_i 的期望值为$\frac{n_i(N+1)}{2}$。

(4) 决策:① 当 $k = 3$ 且每组总体的观测值均不超过5,查附录5,在显著水

平 α 下，若 $T > C_\alpha$，则拒绝 H_0。

② 若无法利用附录 8 时，则查附录 5，在显著水平 α 下，若 $T > \chi^2_\alpha(k-1)$，则拒绝 H_0。

【例 9.7】　哈尔滨某医院妇产科医师，为了研究怀孕妇女其肾上腺素的分泌量变化情况，分别记录三组妇女体内肾上腺素的分泌量。第一组为待产的妇女，第二组为剖腹生产的妇女，第三组为自然生产的妇女，表 9.7 提供了具体记录数据值。在显著水平 $\alpha = 0.05$ 下，试判断这三组妇女肾上腺素的分泌量是否有不同。

表 9.7　三组妇女体内肾上腺素的分泌量反模糊化值

第一组	263	305	218	333	454	339	304	154	287	355
第二组	468	501	455	351	469	362				
第三组	343	772	207	998	838	687				

统计假设为：

H_0：每组的妇女体内肾上腺素的分泌量分布相同；

H_1：至少有一组妇女体内肾上腺素的分泌量分布与其他组不同。

将各组观测值个数之和：$10 + 6 + 6 = 22$ 个观测值加以混合排序之后，我们得到

第一组	4	7	3	8	14	9	6	1	5	12	$R_1 = 69$
第二组	16	18	15	11	17	13					$R_2 = 90$
第三组	10	20	2	22	21	19					$R_3 = 94$

$$K = \frac{12}{22(22+1)}\left[\frac{69^2}{10} + \frac{90^2}{6} + \frac{94^2}{6}\right] - 3(22+1) = 9.23$$

因为由附录 5 知 $\chi^2_{0.05}(2) = 5.99 < 9.232$，所以我们拒绝 H_0，也就是妇女体内肾上腺素的分泌量分布不同。

9.6　结　　论

当研究的问题具有模糊性，而且论域可由数个变量所分割时，由于模糊理论依据受访者的主观意识不同，容许其对于不同的选项作出一个以上的选择，由此所得到的调查结果及统计分析的确较能贴近总体意向的真实情况。这也正

是本章考虑多值逻辑模糊数的排序及检验问题的原因。

我们所提出的模糊样本排序方法，可以应用于非参数检验，用以解决传统假设检验无法解决日常生活中带有模糊性语意，或者带有不确定性区间值的检验假设。应用模糊二元有序关系，以决定不确定性区间值之间的大小排序关系，并应用于非参数检验，结果表明，无论是离散型或是连续型，均较一般模糊数排序方法简单且有效率。

我们所提出的模糊符号检验，模糊威尔克森符号秩和检验，模糊克鲁斯考尔和沃利斯检验等非参数检验方法，具有不需要对数据分布的形式给予明确假设，例如，总体分布是正态或近似正态的优点。此外，当我们想检验针对日常生活常遇到如“我的食量是2～3碗”，“我希望的年薪是3～4万”，“某股票股价是否会在23元左右盘整”，“某大型汽车展今日到展人数是1～2万人”等具有不确定性及模糊性的数据，应视为是一个有关模糊样本的假设检验。传统的统计检验方法对于这样的假设并无有效的解决方法，而我们所提出的模糊非参数检验方法则提供了一个更具体的、易于操作的可行方向。

第10章　模糊数据时间数列分析及预测

本章将把模糊集合理念引入到时间数列分析与预测上，探讨和研究的重点内容包括模糊自相似度的定义与度量，模糊自回归系数的分析，模糊相似度辨识与自回归定阶，模糊数据时间数列模型建立与预测等。首先，本章给出模糊数据时间数列模型的概念和一些重要性质，然后提出模糊相似度的定义与度量，并推导模糊关系方程式，提出模糊数据时间数列模型建立方法，且考虑以平均预测准确度来作预测效果的比较。最后，以中国台湾经济景气对策信号为实例，用以检验所建立的模糊数据时间数列模型及预测的有效性与实用性。

10.1　引　　言

在时间数列分析中，数据的走势形式可以作为判断事件发生的基础，比如递增或递减、季节性循环或突发暴涨等。因此，根据所观察到的特性，可以通过先验的模型族中，例如，由 ARIMA 模型族、ARCH 模型族或门限模型族等，挑选出最佳的拟合模型。但是，由于数据收集的误差、时间的延迟(Lag)或变量之间的交互影响，使得单一度量的数值，在形式上看似一个精确值，而实际上所隐含的却是某一区间范围的可能值。在这种情况下，我们如果仍用传统模型建立与分析方法来拟合出一种数学模型，以解释时间数列数据与走势，可能会产生模型过度拟合的危险。

预测信息是决策者在决策过程中不可或缺的重要信息，且精确的预测结果可以提供决策者更多的信息，有利于作出正确的决策与适当的反应。在时间数列数据上，传统做法是以建立统计模型方法作为数据分析或预测工具，其过程倾向于由数值数据、经由先验的模型族中建立最适合的模型，例如，由 ARIMA 模型族、ARCH 模型族或门限模型族等，来挑选出最佳的拟合模型。但是，由于数据收集的误差、时间的迟延或变量之间的交互影响，使得单一度量的数值，在形式上看似一个精确值，但实际上所表达的应该是某一区间范围的可能值。例如，每年的学生注册人数是以年初、年中、年尾、或是平均为准？期间所得数值往往各有不同。又比如，每天的加权股票指数是以开盘、收盘，或是最高最低价的平均为准？分析结果亦有相当的差距。

实际上，学生注册人数与加权股票指数，其数字本身就具有不确定性与模糊性。倘若我们利用这种假定的精确值来作因果分析或计量度量，可能会造成因果判定偏差或错误的预测模型的建立。因此，具有模糊性的数值数据，如果我们用传统的模型建立与分析，拟合出数学模型，以解释时间数列数据与走势，可能会造成模型错误建立、预测错误或扩大预测结果和实际状态之间的误差。

另一方面，从模型库(Model Base)的观点来看，传统上是以建立统计模型作为数据分析或预测工具。其过程倾向于由数值数据、经由先验的模型族中建立最适合的模型。比如，由 ARIMA 模型族、ARCH 模型族或门限模型族等，来挑选出最佳的拟合模型。但是，由于数据收集的误差、时间的延迟或变量之间的交互影响，数值数据本身经常充满着不确定性。例如，景气对策信号综合指数的判定，有些项目可能实时收录记载，有些项目仅能事后观察得来。

因此，往往动态数据的显现上看似一精确的数值，而所隐含的却是某一区间范围的可能值、或者数据本身就是一组语言数据(Linguistic Data)。这些静态或动态情况，我们如果利用传统的模型建立与分析，拟合出数学模型，以解释时间数列数据与走势，可能会产生模型过度拟合的危险。

把模糊逻辑应用于时间数列分析过程时，第一步要考虑的是，要如何结合语言变量的分析方法，来解决数据的模糊性问题。语言方面提出了具有学习能力的方法去修正模糊模型。还有一些其他方法，比如使用试错法(Trial-and-error Procedure)选择适当的加权因子，但是，此方法相当麻烦。如果是由模糊方程式着手，事实上，与决策表或决策法则相比，模糊关系方程式是容易理解的，且易于更进一步的分析及应用。因此，大部分学者，经常利用模糊关系方程式求解及分析。

本章将详细探讨这些类型的模糊数据时间数列，并用模糊关系方程式建立模糊数据时间数列模型。最后，把我们的方法应用在中国台湾经济景气对策信号数据分析与预测上。

在讨论应用模糊数据时间数列来进行预测时，首先，要给出几个有关模糊数据时间数列的定义。

【定义 10.1】 模糊数据时间数列

设$\{X(t)\in \mathbf{R}, t=1,2,\cdots,n\}$表示一个时间数列，$U$为其论域。给定$U$的一个有序分割集合(Ordered Partition Set，或称为有序分拆集合)，$\{P_i, i=1,2,\cdots,r, \bigcup_{i=1}^{r}P_i=U\}$其相对于语言变量为$\{L_i, i=1,2,\cdots,r\}$。若在$\{L_i, i=1,2,\cdots,r\}$上相对于$X(t)$的模糊集合$F(t)$有隶属函数为$\{\mu_1(X(t)),$

$\mu_2(X(t)),\cdots,\mu_r(X(t))\}$，$0 \leqslant \mu_i(t) \leqslant 1$，$i = 1,\cdots,r$，则称 $F(t)$ 的集合，即 $\{F(t)\}$ 为 $X(t)$ 上的一个模糊数据时间数列，记为

$$F(t) = \frac{\mu_1(X_t)}{L_1} + \frac{\mu_2(X_t)}{L_2} + \cdots + \frac{\mu_r(X_t)}{L_r} \tag{10.1}$$

其中"+"表示连接符号，$\mu_i : \mathbf{R} \to [0,1]$，且 $\sum_{i=1}^{r} \mu_i(X(t)) = 1$，对于所有 $t = 1, 2, \cdots, n$。

为了方便起见，我们把

$$\frac{\mu_1(X_t)}{L_1} + \frac{\mu_2(X_t)}{L_2} + \cdots + \frac{\mu_r(X_t)}{L_r}$$

简记为 $F(t)$。

在计算模糊隶属度时，我们考虑利用三角形隶属函数把时间数列变换成为模糊数据时间数列的程序，其方法如下：

设 $\{X_t \in \mathbf{R}, t = 1,2,\cdots,n\}$ 表示一个时间数列，$\{P_i \mid i = 1,2,\cdots,r, \bigcup_{i=1}^{r} P_i = U\}$ 为论域 U 的一个有序分割集合。设 m_i 为其分割集合的中间值，$i = 1,2,\cdots,r$，且其相对于语言变量为 $\{L_i, i = 1,2,\cdots,r\}$，对任意 t，若 X_t 介于 m_i 与 m_{i+1} 之间，则其属于语言变量 L_i 的隶属度为 $\frac{m_{i+1} - X_t}{m_{i+1} - m_i}$，属于语言变量 L_{i+1} 的隶属度为 $\frac{X_t - m_i}{m_{i+1} - m_i}$，属于其他语言变量 $\{L_i, i = 1,\cdots,i-1,i,i+2,\cdots,r\}$ 的隶属度皆为 0。

【例 10.1】　设 $\{X(t)\} = \{0.8,1.7,2.9,4.1,3.5,3.2,4.3,3.6\}$。选取 $U = \{[0,1],[1,2],[2,3],[3,4],[4,5]\}$，且其语言变量为：很低 $= L_1 \propto [0,1]$，低 $= L_2 \propto [1,2]$，中等 $= L_3 \propto [2,3]$，高 $= L_4 \propto [3,4]$，很高 $= L_5 \propto [4,5]$，其中"$\propto$"表示"相对于"。同时，我们取区间的平均数为典型值。因此，各个语言变量相对应的典型值分别为 0.5,1.5,2.5,3.5,4.5，其隶属函数如图 10.1 所示。

于是，我们可得到相对于 $\{X(t)\}$ 的模糊数据时间数列 $\{F(t)\}$ 为

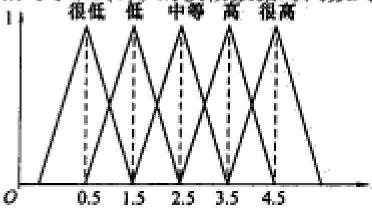


图 10.1　$\{F(t)\}$ 相对应的隶属函数

	很低	低	中等	高	很高	
$F(1)=($	0.7	0.3	0	0	0	$)$
$F(2)=($	0	0.8	0.2	0	0	$)$
$F(3)=($	0	0	0.6	0.4	0	$)$
$F(4)=($	0	0	0	0.4	0.6	$)$
$F(5)=($	0	0	0	1	0	$)$
$F(6)=($	0	0	0.2	0.8	0	$)$
$F(7)=($	0	0	0	0.2	0.8	$)$
$F(8)=($	0	0	0	0.9	0.1	$)$

在时间数列分析与模型建立中，自相关值的决定相当重要，因为它反映出一个模糊数据时间数列过程的自相关趋势。但是，对一组不确定或不完善数据而言，自相关应不可以由一个数值来说明。因此，本章研究考虑用模糊关系来分析它们的自相关程度。对模糊数据时间数列而言，模糊关系及模糊关系矩阵，也是一个主要探讨的重点。

下面，我们给出模糊数据时间序列经常用到的模糊关系及模糊关系矩阵定义。

【定义 10.2】 模糊关系(Fuzzy Relation)

设$\{P_1,P_2,\cdots,P_r\}$表示论域 U 的一个有序分割。若在 U 上的模糊集 F 与 G 分别为 $F=(\mu_1,\cdots,\mu_r)$ 与 $G=(v_1,\cdots,v_r)$，其中 μ_i 与 v_j 表示 U 上的隶属度，则介于 F 与 G 之间的模糊关系记为

$$R=F^{\mathrm{T}}\circ G=[R_{ij}]_{m\times m} \tag{10.2}$$

其中"$\circ$"表示模糊合成运算，在此我们使用最大－最小(max－min)，"T"表示转置，R_{ij} 表示介于 F 与 G 之间关系的隶属度。

[评注]

(1) 设 $R(t,t-1)$ 是 1 阶的模糊关系矩阵。若对任何时点 t，$R(t,t-1)$ 与 t 无关，亦即对任何 t，$R(t,t-1)=R(t-1,t-2)$，则称 $F(t)$ 具有时常值的(Time-invariant)性质，否则称为具有时变值的(Time-variant)性质。

(2) 设 $F(t)$ 表示一个 $FAR(1)$ 的时常的模糊数据时间数列。若对任何时点 t，恒有 $F(t)=F(t-1)$ 且 $F(t)$ 是由有限个 $f_{ik}(t)$，$k=1,2,\cdots,m$ 所组成，则

$$R(t,t-1)=f_{i1}(t-1)\wedge f_{i1}(t)\vee f_{i2}(t-2)\wedge f_{i2}(t-1)\vee\cdots\vee f_{im}(t-m)\wedge f_{jm}(t-m+1)$$

其中 i,j 为对应的语言程度指标(Index of Linguistic Hedge)，$m>0$，$\wedge$ 与 $\vee$ 分别表示最小与最大的运算。

通常，要了解和认识两模糊时间数列的关系程度，首先考虑其相似性。由于相似度与两时间数列间的距离有关，其大小应与距离成反比。此外，时间数列属于动态数据，两时间数列间的关系存在的时间落差，具有很好的参考价值。

【定义 10.3】　两模糊数据时间数列的距离

设 $\{FX_t = (mx_{t1}, mx_{t2}, \cdots, mx_{tK}), t = 1, \cdots, w\}$，$\{FY_t = (my_{t1}, my_{t2}, \cdots, my_{tK}), t = 1, \cdots, w\}$ 分别表示由两模糊数据时间数列 $\{X_t \mid t = 1, 2, \cdots, m\}$，$\{Y_t \mid t = 1, 2, \cdots, n\}$ 各抽取长度为 w 的模糊时间数列子集合，则把 $\{FX_t\}$，$\{FY_t\}$ 两时间数列的距离定义为

$$d(FX_t, FY_t) = \left| \sum_{k=1}^{K} (mx_{tk} - my_{tk}) \cdot j \right| \tag{10.3}$$

【例 10.2】　已知两模糊数据时间数列 $\{FX_t\}$ 与 $\{FY_t\}$ 值，如表 10.1 所示。

表 10.1　模糊数据时间数列 $\{FX_t\}$ 与 $\{FY_t\}$ 的值

	L_1	L_2	L_3	L_4	L_5		L_1	L_2	L_3	L_4	L_5
FX_1	0	0	1	0	0	FY_1	0.5	0	0.5	0	0
FX_2	0.3	0.7	0	0	0	FY_2	0.4	0.6	0	0	0
FX_3	0	0.4	0.6	0	0	FY_3	0	0.8	0.2	0	0
FX_4	0	0	0.9	0.1	0	FY_4	0	0	0.5	0.5	0
FX_5	0	0	0	1	0	FY_5	0	0	0.5	0	0.5

则表 10.2(a) 至 10.2(e) 分别表示模糊数据时间数列 $\{FX_t\}$ 与 $\{FY_t\}$ 为 $w = 1$ 至 5 期的绝对距离值。

表 10.2　(a) 模糊数据时间数列 $\{FX_t\}$ 与 $\{FY_t\}$ 为 $w = 1$ 期的距离值

d	FY_1	FY_2	FY_3	FY_4	FY_5
FX_1	1	1.4	0.8	0.5	1
FX_2	0.3	0.1	0.5	1.8	2.3
FX_3	0.6	1	0.4	0.9	1.4
FX_4	1.1	1.5	0.9	0.4	0.9
FX_5	2	2.4	1.8	0.5	0

(b) 模糊数据时间数列$\{FX_t\}$与$\{FY_t\}$为 $w=2$ 期的距离值

d	FY_1,FY_2	FY_2,FY_3	FY_3,FY_4	FY_4,FY_5
FX_1,FX_2	2.8	2.8	3.6	5.6
FX_2,FX_3	2	2	3.6	6.4
FX_3,FX_4	4.2	3.8	2.6	3.6
FX_4,FX_5	7	6.6	3.6	1.8

(c) 模糊数据时间数列$\{FX_t\}$与$\{FY_t\}$为 $w=3$ 期的距离值

d	FY_1,FY_2,FY_3	FY_2,FY_3,FY_4	FY_3,FY_4,FY_5
FX_1,FX_2,FX_3	6.1	7.4	9.6
FX_2,FX_3,FX_4	6.4	7.5	9.5
FX_3,FX_4,FX_5	11.7	9.8	7.2

(d) 模糊数据时间数列$\{FX_t\}$与$\{FY_t\}$为 $w=4$ 期的距离值

d	FY_1,FY_2,FY_3,FY_4	FY_2,FY_3,FY_4,FY_5
FX_1,FX_2,FX_3,FX_4	9.6	10.2
FX_2,FX_3,FX_4,FX_5	12.6	12.2

(e) 模糊数据时间数列$\{FX_t\}$与$\{FY_t\}$为 $w=5$ 期的距离值

d	FY_1,FY_2,FY_3,FY_4,FY_5
FX_1,FX_2,FX_3,FX_4,FX_5	25.5

此距离代表的含义，以表 10.2(a) 为例，$d(FX_5,FY_5)$ 的值最小，代表 FX_5，FY_5 这两个模糊数据时间数列值最接近（相似）；而 $d(FX_5,FY_2)$ 的值最大，代表 FX_5，FY_2 这两个模糊数据时间数列的值最不接近（不相似）。同样地，在表 10.2(c) 中，我们发现，(FX_1,FX_2,FX_3) 与 (FY_1,FY_2,FY_3) 最相似，(FX_3,FX_4,FX_5) 与 (FY_1,FY_2,FY_3) 最不相似。

10.2 模糊数据时间数列的统计分析

在建立模糊数据时间数列模型之前，我们需要了解模糊趋势，模糊关系的计算方法，以及对模糊数据时间数列阶数进行确定等问题。

模糊趋势(Fuzzy Trend)

对于大多数时间数列数据而言，均呈现非稳定的(Non-stationary)状态，因而，在分析数据之前，需要使得数据达到稳定(Stationary)状态，以便于分析。因此，对于模糊数据时间数列而言，为了达到更精确的结果，也需要先使数据达到稳定状态，然后再进行更进一步的分析。然而，如何去判断模糊时间数列是否稳定呢?事实上，并没有一个固定的方法，而我们提出下列的定义与决策法则，来检验模糊趋势。

【定义 10.4】　模糊高于(Fuzzy Higher)

设 $\{F(t), t = 1,2,\cdots,n\}$ 表示论域 U 上的模糊时间数列。设 $F(i)$ 与 $F(j)$ 分别有最大隶属度 $\mu_r(i)$，$\mu_s(j)$ 分别落于有序分割集合 P_r 与 P_s。若(1)P_r 高于 P_s，或(2)$P_r = P_s$，且 $\mu_{r+1}(i) > \mu_{s+1}(j)$，则称 $F(i)$ 是模糊高于 $F(j)$，记为 $F(i) >_F F(j)$。若 $P_r = P_s$ 且 $\mu_r(i) = \mu_s(j)$，则称 $F(i)$ 与 $F(j)$ 模糊相等(Fuzzy Equal)，记为 $F(i) =_F F(j)$。

【例 10.3】　把天气气象情况用模糊集来表示是合理的，因为天气本身是一种语言变量。因此，天气状况可以用模糊时间数列表示。设 $F(1)$，$F(2)$ 以及 $F(3)$ 在模糊时间数列上的模糊集表示“热”、“温暖”以及“有点热”，则 $F(1)$，$F(2)$ 以及 $F(3)$ 可表为

$$F(1) = \frac{0.0}{[20,25)} + \frac{0.2}{[25,30)} + \frac{0.8}{[30,35]}$$

$$F(2) = \frac{0.0}{[20,25)} + \frac{0.8}{[25,30)} + \frac{0.2}{[30,35]}$$

$$F(3) = \frac{0.0}{[20,25)} + \frac{0.6}{[25,30)} + \frac{0.4}{[30,35]}$$

其中，论域 $U = [20,25) \cup [25,30) \cup [30,35]$ 表示温度的范围。

(1) 因为 $F(1)$ 及 $F(2)$ 分别有最大的隶属度 0.8 在(30,35] 及(25,30]，根据定义 10.3，所以 $F(1) >_F F(2)$。

(2) 因为 $F(2)$ 及 $F(3)$ 分别有最大的隶属度都在(20,30]，又 $F(2)$ 及 $F(3)$ 在(30,35) 的隶属函数值分别为 0.2 及 0.4，根据定义 10.3，所以 $F(3) >_F F(2)$。

(3) 显然，$F(1) >_F F(3)$。

在考虑模糊时间数列高低关系时，我们需要用到的模糊数据时间数列高低关系的概率测度，其概念可如下定义。设

$$P(F(t) >_F F(t-1)) = \frac{(F(t) >_F F(t-1)\text{ 的个数})}{[n-1-(F(t) =_F F(t-1)\text{ 的个数})]} \tag{10.4}$$

【定义 10.5】 模糊趋势

设$\{F(t), t = 1,2,\cdots,n\}$表示一个模糊数据时间数列。对于任一时点$t > 1$,若

$$P(F(t) >_F F(t-1)) \geqslant 1-\lambda \quad 0 \leqslant \lambda < 0.5 \tag{10.5}$$

则称$\{F(t)\}$在显著水平λ上有模糊向上趋势(Fuzzy Upward Trend)。另一方面,若

$$P(F(t-1) >_F F(t)) \geqslant 1-\lambda \quad 0 \leqslant \lambda < 0.5 \tag{10.6}$$

则称$\{F(t)\}$在显著水平λ上有模糊向下趋势(Fuzzy Downward Trend)。在式(10.5)中,若$\lambda = 0$称$\{F(t)\}$是绝对向上趋势(Absolutely Upward Trend);而在式(10.6)中,若$\lambda = 0$称$\{F(t)\}$是绝对向下趋势(Absolutely Downward Trend)。若$P(F(t) >_F F(t-1)) = P(F(t-1) >_F F(t)) = 0.5$,则称$\{F(t)\}$是没有趋势。

模糊趋势检验法则

(1) 建立统计假设检验。

H_0:没有模糊向上或向下趋势;

H_1:具有模糊向上或向下趋势。

(2) 设定显著水平λ以及检验统计量

$$P(F(t-1) >_F F(t)) \text{或} P(F(t-1) >_F F(t))$$

(3) 应用单侧检验,在显著水平λ下,若$P(F(t-1) >_F F(t)) > 1-\lambda$,或$P(F(t-1) >_F F(t)) > 1-\lambda$,则拒绝$H_0$。

10.3 模糊数据自回归时间数列($FAR(p)$)

首先,假设一个多输入单一输出(Multiple Input Single Output,MISO)的多变量模糊数据时间数列(p阶)如下

$$\begin{aligned} F(t) = {} & F(t-1) \circ F(t-2) \circ \cdots \circ F(t-p) \circ \\ & I_1(t) \circ I_1(t-1) \circ \cdots \circ I_1(t-q_1) \circ \cdots \circ \\ & I_k(t) \circ I_k(t-1) \circ \cdots \circ I_k(t-q_k) \circ R \end{aligned} \tag{10.7}$$

其中$t = 1,2,3,\cdots$表示抽样时点,$F(\cdot)$表示模糊时间数列,$I_i(\cdot), i = 1,2,\cdots,k$表示隶属度输入值,$R$表示$I_i$与$F$的模糊关系矩阵,$p$表示模糊时间数列阶数(Order),$q_1, q_2, \cdots, q_k$表示相对此模糊时间数列的模糊隶属度值。

为了方便起见,设$G_1(t) = F(t-1), \cdots, G_p(t) = F(t-p), G_{p+1}(t) =$

$I_1(t),\cdots,G_n(t)=I_k(t-q_k)$，则式(10.5)的模糊数据时间数列可以改写为

$$F(t)=G_1(t)\circ G_2(t)\circ\cdots\circ G_n(t)\circ R \tag{10.8}$$

其中 $n=p+q_1+\cdots+q_k$。设 U_i 及 V 分别表示模糊输入 G_i 及模糊输出 F 的论域，并且每一个论域具有相同个数 r 的分割模糊集，记为 $M=\{1,2,\cdots,r\}$。则我们可以将观测值 $x_i(t)\in U_i$ 及 $y(t)\in V$ 与模糊集 $G_i(t)$ 及 $F(t)$ 结合如下

$$G_i(t)=\{A_j^i(x_i(t))\mid j\in M\} \tag{10.9}$$

$$F(t)=\{B_j(y(t))\mid j\in M\} \tag{10.10}$$

其中 A_j^i 及 B_j 分别表示模糊集 $G_i(t)$ 及 $F(t)$ 的隶属函数。这些隶属函数，一般定义为

$$A_j^i:U_i\to[0,1],j\in M,i=1,2,\cdots,k \tag{10.11}$$

$$B_j:V\to[0,1],j\in M \tag{10.12}$$

为了方便起见，以后隶属函数值 $A_j^i(x_i(t))$ 及 $B_j(y(t))$ 分别用 $a_j^i(t)$ 及 $b_j(t)$ 表示。因此，每一数据均为有 r 个元素的隶属函数值的集合，且模糊关系矩阵可以表为 $\boldsymbol{R}=\boldsymbol{I}_1\times\boldsymbol{I}_2\times\cdots\times\boldsymbol{I}_n\times\boldsymbol{F}$，其中"×"表示笛卡儿乘积(Cartesians Product)。若隶属函数 A_j^i 及 B_j 有其语言意义，比如"大"，"好"…，则 $\boldsymbol{R}$ 中的每一元素表示为 $R_{j_1,j_2,\cdots,j_n,j}$，其相对应的语言规则为

$$\text{若 } x_1 \text{ 是 } A_{j_1}^1,\cdots,x_n \text{ 是 } A_{j_n}^n,\text{则 } y \text{ 是 } B_j \tag{10.13}$$

因此，模糊关系方程式可以被视为语言的模糊规则。

在探讨模糊数据时间数列的类型之前，首先要把数据转换为模糊数据，而本章我们利用 Z 型、S 型以及 π 型三种隶属函数把观测值转换为模糊集合，如图 10.2 所示。

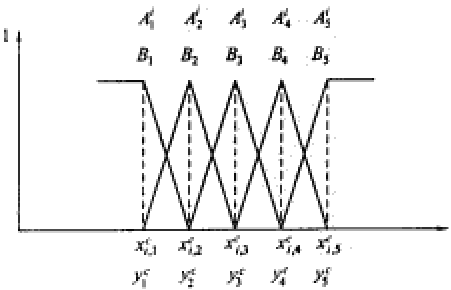


图 10.2　$\{F(t)\}$ 与 $\{G(t)\}$ 相对应的隶属函数

因此，观测值 $x_i(t)$ 及 $y_j(t)$ 分别可以用隶属函数值 $a_j^i(t)$ 及 $b_j(t)$ 表示如下

$$x_i(t) = \sum_{j=1}^{r} x_{i,j}^c a_j^i(t) \tag{10.14}$$

$$y(t) = \sum_{j=1}^{r} y_j^c b_j(t) \tag{10.15}$$

其中，$x_{i,j}^c$ 表示论域 U_i 的第 j 个分割的典型值，y_j^c 表示论域 V 的第 j 个分割的典型值，A_j^i 表示第 i 组输入数据的第 j 个隶属函数，B_j 表示输出数据的第 j 个隶属函数。

若上述关系得到满足，则可以得出下面性质。

【性质 10.1】 设隶属函数是标准隶属函数，则观测值与隶属函数值是一一对应且映上的(One-to-one Correspondence and Onto)。

证明 因为 $x_i(t) = \sum_{j=1}^{r} x_{i,j}^c$，且 $\sum a_j^i(t) = 1, \forall i = 1,2,\cdots,n$。又 $a_j^i(t)$，$b_j(t)$ 均为隶属函数值，故存在 $a_k^i(t), a_{k+1}^i(t), \forall t, k \in \{1,2,\cdots,r-1\}$，使得 $a_k^i(t) + a_{k+1}^i(t) = 1$ 且 $a_j^i(t) = 0, \forall j \neq k, k+1$。因而

$$x_i(t) = x_{i,k}^c a_k^i(t) + x_{i,k+1}^c a_{k+1}^i(t) = x_{i,k}^c a_k^i(t) + x_{i,k+1}^c(1 - a_k^i(t)) = (x_{i,k}^c - x_{i,k+1}^c) a_k^i(t) + x_{i,k+1}^c$$

因此，$x_i(t)$ 与 $a_k^i(t)$ 为一一对应且映上的。同理，$y(t)$ 与 $b_k(t)$ 亦为一一对应且映上的。

一旦具有上面的说明，我们就可分析模糊数据时间数列模型。在此，为了便于阐述思想，仅仅探讨单变量自回归模糊数据时间数列。下面是一些常见的模糊数据时间数列模型。

1 阶自回归模糊数据时间数列 $FAR(1)$：若

$$F(t) = F(t-1) \cdot R \tag{10.16}$$

则称其阶数为 1 的模糊自回归模型(First Order of Fuzzy Auto-regressive Model)，记为 $FAR(1)$，$\{F(t), t = 1,\cdots,n\}$ 表示数据 $\{X(t), t = 1,\cdots,n\}$ 所对应语言变量的隶属度集合，R 为 $F(t)$ 与 $F(t-1)$ 之间的模糊关系矩阵。

p 阶自回归模糊时间数列 $FAR(p)$：若

$$F(t) = F(t-1) \cdot F(t-2) \cdot \cdots \cdot F(t-p) \cdot R \tag{10.17}$$

称其阶数为 p 的模糊自回归模型(p-th Order of Fuzzy Auto-regressive Model)，R 为 $F(t)$ 与 $F(t-1), F(t-2), \cdots, F(t-p)$ 之间的模糊关系矩阵。

对于季节性数据，可以考虑下述模型。

1 阶季节性自回归模糊时间数列 $FSAR(1)_s$：若

$$F(t) = F(t-s) \cdot R \tag{10.18}$$

其中 s 表示季节周期，$\{F(t), t = 1,\cdots,n\}$ 表示数据 $\{X(t), t = 1,\cdots,n\}$ 所对

应语言变量的隶属度集合，R 为 $F(t)$ 与 $F(t-s)$ 之间的模糊关系矩阵。

p 阶季节性自回归模糊时间数列 $FSAR(p)_s$：若

$$F(t)=F(t-s)\circ F(t-2s)\circ\cdots\circ F(t-ps)\circ R \tag{10.19}$$

称其为 p 阶季节性模糊自回归模型。

10.4　模糊数据时间数列模型建立

在模糊数据时间数列模型的建立过程中，有许多的情况需要注意，现在分别叙述如下：

由于所收集的数据其形式可能为数值的，也可能是定性数据，又有可能是语言值(Linguistic Values)的数据，例如，通过品尝所获得的数据等。对于这些数据，很难只使用传统的时间数列方法来进行分析，若能使用模糊集合的方法，则不会受到数据形式的影响，而且更能建立适合的模型。

而在创立模糊论域集合时，其分割的集合数到底要多少个，目前尚无定论。一般地说，分割集合数越多，则其准确度越高，而相对的复杂度也增加。因此，准确度与复杂度的取舍就要视个人的需要而定。

在数据的模糊化方面，我们是利用标准的隶属函数来把它模糊化。而在此过程中，其典型值的取得尚未有固定方法，我们认为，可以用每个集合中元素的中位数、平均数以及分类的聚类中心为其典型值。但是，若以群聚中心为其典型值，则易造成一个集合中有多个典型值，此时该集合为 Π 型隶属函数，将会使得过程较为复杂，若使用平均数或中位数为典型值，则每一个集合只有一个典型值，所以此时集合为一个 Λ 型隶属函数，问题将会变得较为简单。

有关模糊趋势的概念是由吴柏林等(1994)首先提出的。一般而言，若遇到的数据为数值数据转换为模糊数据，而此时数据具有模糊趋势，其处理的方法是将原始数据差分，使之达到平稳，然后再继续前述的步骤，来建立模型。但是，对于原本即为模糊数据，若能检查出具有模糊趋势，则目前尚未有方法来解决，这仍旧是一个值得继续探讨的问题。

在模糊时间数列的分析上，对时间序列的定阶占据着相当重要的地位。倘若我们能找出正确的阶数，则才能知道真正影响的因子，对于模型的建立才能真正有把握。然而，在以往的文献中，比如 Pedrycz(1984)，宋和奇斯瑟姆(Song and Chissom，1993)，Chen(1996)，均是强迫给定其阶数，没有一个固定的法则，以至于在模型的建立上并不能得到一个正确的模型，因此，预测的结果当然不理想。或者是，尝试各种阶数的组合，比如 Xu(1989)，Lee 等(1994)，这无疑将增加计算上的负担。而本章利用模糊自相似度的概念，来决定其阶数的选取。

模糊数据时间数列的定阶

对于模糊数据时间数列进行定阶来说，到目前为止尚未有一个行之有效的方法。在以往的文献中都是直接给定的。比如，Pedrycz(1984)，宋和奇斯瑟姆(Song and Chissom，1993)，Lee 等(1994)。但是，这一方法太过于主观，理论基础亦薄弱。因此，预测结果并不理想。而本章的研究尝试以模糊自相似度(Fuzzy Auto-similarity，记为 FAS) 的方法，来确定模糊时间数列的阶数。

在探讨模糊自相似度之前，首先必须要了解模糊相似度(Fuzzy Similarity，简记为 FS) 的概念，然后才能把它推广到模糊时间数列的认定。对于模糊相似度的测度方法，我们分别叙述如下。

值域均分法

(1) 已知两时间数列 $\{X(t)\}$，$\{Y(t)\}$，$t = 1,2,\cdots,n$。

设 $R_X = \max(X(t)) - \min(X(t))$，$R_Y = \max(Y(t)) - \min(Y(t))$。

(2) 将两时间数列各分为 k 类，令

$$R_{X_i} = [\min(X(t)) + \frac{i-1}{k}R_X, \min(X(t)) + \frac{i}{k}R_X), i = 1,2,\cdots,k-1$$

$$R_{X_k} = [\min(X(t)) + \frac{k-1}{k}R_X, \max(X(t))]$$

$$R_{Y_j} = [\min(Y(t)) + \frac{j-1}{k}R_Y, \min(Y(t)) + \frac{j}{k}R_Y), i = 1,2,\cdots,k-1$$

$$R_{Y_k} = [\min(Y(t)) + \frac{k-1}{k}R_Y, \max(Y(t))]$$

其中 R_{X_i}，R_{Y_j} 代表 $\{X(t)\}$，$\{Y(t)\}$ 各自将值域 k 等分后所得的 k 个区间的第 i 及 j 个区间，$i,j = 1,2,3,\cdots,k$。

(3) 用 S 表示"$\{X(t)\}$，$\{Y(t)\}$ 两时间数列之间相似"的模糊二元关系集合。我们把两时间数列的配对时点 $(X(t), Y(t))$ 间的汉明距离，作为 $(X(t), Y(t)) \in S$ 的隶属函数的建立基础，则称

$$\mu_S(X(t), Y(t)) = \begin{cases} 1, X(t) \in R_{X_i}, Y(t) \in R_{Y_j}, 且\ i = j \\ \left(\frac{k-2}{k-1}\right)^r, X(t) \in R_{X_i}, Y(t) \in R_{Y_j}, 且\ |i-j| = 1 \\ \vdots \\ \left(\frac{1}{k-1}\right)^r, X(t) \in R_{X_i}, Y(t) \in R_{Y_j}, 且\ |i-j| = k-2 \\ 0, X(t) \in R_{X_i}, Y(t) \in R_{Y_j}, 且\ |i-j| = k-1 \end{cases} \quad (10.20)$$

为$(X(t),Y(t))$属于 S 的隶属函数,其中 $i,j=1,2,\cdots,k$,而 v 表示调整因子,一般选择 $v=1,2$。

若时间数列其方差有所改变时,可考虑以模糊 c 均值(Fuzzy c-means)值域均分法来建立隶属函数。

模糊 c 均值值域均分法

(1) 首先利用 c 均值(c-means)法,分别找出两时间数列$\{X(t)\}$,$\{Y(t)\}$的 k 个聚类中心

$$C_X=\{C_{X_i}\mid i=1,2,\cdots,k \text{ 且 } C_{X_1}<C_{X_2}<\cdots<C_{X_k}\}$$

$$C_Y=\{C_{Y_j}\mid j=1,2,\cdots,k \text{ 且 } C_{Y_1}<C_{Y_2}<\cdots<C_{Y_k}\}$$

(2) 将两时间数列各分为 k 类,令

$$R_{X_1}=[\min(X(t)),\frac{C_{X_1}+C_{X_2}}{2})$$

$$R_{X_i}=[\frac{C_{X_{i-1}}+C_{X_i}}{2},\frac{C_{X_i}+C_{X_{i+1}}}{2}),i=2,3,\cdots,k-1$$

$$R_{X_k}=[\frac{C_{X_{k-1}}+C_{X_k}}{2},\max(X(t))]$$

$$R_{Y_1}=[\min(Y(t)),\frac{C_{Y_1}+C_{Y_2}}{2})$$

$$R_{Y_i}=[\frac{C_{Y_{i-1}}+C_{Y_i}}{2},\frac{C_{Y_i}+C_{Y_{i+1}}}{2}),i=2,3,\cdots,k-1$$

$$R_{Y_k}=[\frac{C_{Y_{k-1}}+C_{Y_k}}{2},\max(Y(t))]$$

其中 R_{X_i},R_{Y_j} 代表$\{X(t)\}$,$\{Y(t)\}$中各自将值域 k 等分后所得的 k 个区间。

(3) 同值域均分法步骤(3) 建立隶属函数 $\mu_S(X(t),Y(t))$。

当时间数列数据有离群值或突发值时,为了消除离群值或突发值的干扰,我们可利用序关系(Order Relation)变换法将原始数据加以排序,再进行建立隶属函数的工作。

序关系变换法

(1) 先将两时间数列$\{X(t)\}$,$\{Y(t)\}$分别排序可得其有序数列为:$\{R_X\mid t=1,2,\cdots,n\}$,$\{R_Y\mid t=1,2,\cdots,n\}$。

(2) 用 S 表示"$\{X(t)\}$,$\{Y(t)\}$两时间数列相似"的模糊二元关系集合。我们将两时间数列的配对时点$(X(t),Y(t))$间的汉明距离,作为$(X(t),

$Y(t)) \in S$ 的隶属函数的建立基础，则我们称

$$\mu_S(X(t),Y(t)) = \left(1 - \frac{|R_{X(t)} - R_{Y(t)}|}{n-1}\right)^v$$

为$(X(t),Y(t)) \in S$的隶属函数，其中 v 表示调整因子(为了方便起见，通常 v 取正整数)。

一旦具有上述概念，我们可以定义模糊相似度如下。

【定义 10.6】 两个时间数列之间的模糊相似度

已知两时间数列$\{X(t)\}$，$\{Y(t)\}$，$t=1,2,\cdots,n$。若 U，V分别表示$X(t)$及$Y(t)$的论域，S表示"$\{X(t)\}$，$\{Y(t)\}$两时间数列相似"的模糊二元关系集合，隶属函数为$\mu_S(X(t),Y(t))$，则称

$$FS_{XY} = \frac{1}{n}\sum_{t=1}^{n}\mu_S(X(t),Y(t)) \tag{10.21}$$

为两时间数列之间的模糊相似度。若 FS_{XY} 越大者表示两时间数列其模糊相似度越高，其中$0 \leqslant FS_{XY} \leqslant 1$，且$FS_{XY}=1$表示两时间数列之间完全相似，$FS_{XY}=0$表示两时间数列之间完全不相似。

【定义 10.7】 模糊相似性大小

已知三个时间数列$\{X(t)\}$，$\{Y(t)\}$，$\{Z(t)\}$，$t=1,2,\cdots,n$。若时间数列$\{X(t)\}$比时间数列$\{Z(t)\}$更相似于时间数列$\{Y(t)\}$，则称FS_{XY}大于FS_{ZY}，记作$FS_{XY} > FS_{ZY}$。

【例 10.4】 已知三个时间数列

$$\{X(t)\} = \{0.8,1.9,2.7,4.1,3.6,2.9,4.4,3.7\}$$
$$\{Y(t)\} = \{17,20,22,39,35,28,44,29\}$$
$$\{Z(t)\} = \{33,45,28,97,22,26,49,53\}$$

其走势图如图 10.3 所示。

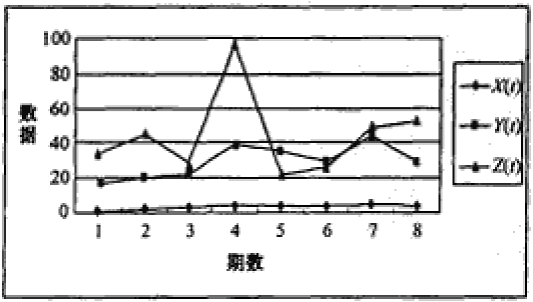


图 10.3 $\{X(t)\}$，$\{Y(t)\}$，$\{Z(t)\}$的走势图

(1) 若采用值域均分法可得$\{X(t)\}$，$\{Y(t)\}$两时间数列之间相似的隶属

度集合为 $\mu_S(X(t),Y(t))=\{1,0.75,0.5,1,1,1,1,0.5\}$，两时间数列之间的模糊相似度 $FS_{XY}=\frac{1}{8}\sum_{t=1}^{8}\mu_S(X(t),Y(t))=0.84$。而 $\{Y(t)\}$，$\{Z(t)\}$ 相似的隶属度集合为，$\mu_S(Z(t),Y(t))=\{1,0.75,1,1,0.25,0.5,0.5,1\}$，两时间数列之间的模糊相似度 $FS_{ZY}=\frac{1}{8}\sum_{t=1}^{8}\mu_S(Z(t),Y(t))=0.75$。

(2) 类似地，若采用模糊 c 均值值域均分法或秩转换法，亦可得到 FS_{XY} 及 FS_{ZY}，我们将此三法所得结果整理如表 10.3 所示。

表 10.3　三种不同方法的模糊相似度评比

	FS_{XY}	FS_{ZY}
值域均分法	0.84	0.75
模糊 c 均值法	0.94	0.69
序关系变换法	0.96	0.68

结果表明，不论是利用何种方法建立隶属函数，其结果均是 $\beta_{XY}>\beta_{ZY}$，即时间数列 $\{X(t)\}$ 比 $\{Z(t)\}$ 更相似于时间数列 $\{Y(t)\}$。

一旦具有模糊相似度以及模糊相似性大小的概念之后，我们可以将其推广到模糊自相似度的概念，并以此方法来决定模糊时间数列的阶数。

【定义 10.8】　模糊自相似度

设 $\{X(t),t=1,2,\cdots,n\}$ 表示一个时间数列，若 U 分别表示 $X(t)$ 的论域，其相对应的模糊时间数列为 $\{F(t),t=1,2,\cdots,n\}$，S 表示"$\{F(t)\}$ 落后 l 期自相似" 的模糊二元关系集合，隶属函数为 $\mu_S(F(t),F(t-l))$。则我们称

$$FAS_l=\frac{1}{n-l}\sum_{t=l+1}^{n}\mu_S(F(t),F(t-l)) \tag{10.22}$$

为模糊时间数列 $\{F(t)\}$ 落后 l 期的模糊自相似度。

为了更符合实际问题的使用，我们给出调整后的模糊自相似度（Adjust Fuzzy Auto-similarity，记为 AFAS）的定义如下。

【定义 10.9】　调整后的模糊自相似度

设 FAS_l，$l=1,2,\cdots,n-1$，表示 $\{F(t)\}$ 落后 l 期的模糊自相似度，k 为选定的时期数，则把调整后的模糊自相似度定义为

$$AFAS_l=\frac{FAS_l-\min\limits_{1\le l\le k}\{FAS_l\}}{1-\min\limits_{1\le l\le k}\{FAS_l\}} \tag{10.23}$$

【定义 10.10】　p 阶数的模糊时间数列

设 $\{X(t),t=1,2,\cdots,n\}$ 表示一个时间数列，其相对应的模糊数据时间数

列为$\{F(t), t=1,2,\cdots,n\}$，α为显著水平，若$AFAS_1, AFAS_2, \cdots, AFAS_p$皆大于显著水平$\lambda$（这里指水平集($\lambda$ – level)），则

$$\{F(t)\} \text{ 与 } \{F(t-1)\}, \{F(t-2)\}, \cdots, \{F(t-p)\}$$

有显著相似。我们称此模糊时间数列为阶数p的模糊时间数列，记为$FAR(p)$。

模糊关系矩阵 *R* 的计算

由10.2节中的模糊数据时间数列模型分析可知，模糊关系矩阵R是建立模型好坏的关键因素。若能准确地掌握模糊关系矩阵R，则模型将会有较佳的拟合。模糊关系矩阵R的计算有许多不同的方法。但是都是基于不同的论点而进行。下面是针对模糊数据时间数列所作的考虑。

假定$F(t)(t=1,2,\cdots,n)$是一个模糊时间数列，$i,j=\{1,2,\cdots,r\}$表示模糊数据时间数列分类的指标集合。为了方便起见，我们直接用$f_j(t)$及$f_j(t-1)$分别表示模糊集合$f_j(t)$及$f_j(t-1)$的隶属函数。

下面是各种不同的模糊数据时间数列的模糊关系矩阵R的计算方法。

当选定可能的模糊时间数列的阶数之后，利用10.3节所定义的计算方法计算模糊关系矩阵。虽然李(Lee,1994)曾提出利用实际的模糊输出与估计的模糊输出，可以得到较小的误差，但是这种计算方法牵涉太多的数值运算，又过于复杂繁琐。因此，本文仍采用吴柏林(1986)以及Sugeno和Tanaka(1991)提出的方法来决定模糊关系矩阵R。

FAR(1)

设$F(t)$表示一个$FAR(1)$的时常值的模糊数据时间数列。对于任意$f_j \in F(t), j \in J$，存在$f_i \in F(t-1), i \in I$，以及模糊关系矩阵R，使得$f_j(t) = f_i(t-1) \circ R_{ij}$，且$R = \bigcup_{i,j} R_{ij}$，因为$f_j(t) = f_i(t-1) \circ R_{ij}$是等价于“若$f_i(t-1)$则$f_j(t)$”有“$R_{ij} = f_i(t-1) \otimes f_j(t)$”，则

$$R = \sup_{2\leqslant t\leqslant n} \{\min_{i,j}(f_i(t-1), f_j(t))\} \tag{10.24}$$

FAR(*p*)

设$F(t)$表示一个$FAR(p)$的时常值的模糊数据时间数列。对于任意$f_j \in F(t), j \in J$，存在$f_{i_k} \in F(t-k), i_k \in I$，以及模糊关系矩阵$R$，使得

$f_j(t) = (f_{i_1}(t-1) \circ \cdots \circ f_{i_p}(t-p)) \circ R_{i_1,i_2,\cdots,i_p,j}$，且$R = \bigcup_{i_1,i_2,\cdots,i_p,j} R_{i_1,i_2,\cdots,i_p,j}$

因为

$$f_j(t) = (f_{i_1}(t-1) \circ f_{i_2}(t-2) \circ \cdots \circ f_{i_p}(t-p))$$

$R_{i_1,i_2,\cdots,i_p,j}$是等价于“若$f_{i_1}(t-1)$ or $f_{i_2}(t-2)$ or $\cdots$ or $f_{i_p}(t-p)$，则$f_j(t)$”，则有

$$R_{i_1,i_2,\cdots,i_p,j} = \min_j(\max_{i_1,i_2,\cdots,i_p}(f_{i_1}(t-1), f_{i_2}(t-2), \cdots, f_{i_p}(t-p)), f_j(t))$$

因而

$$R = \sup_{p+1 \leq l \leq n} R_{i_1, i_2, \cdots, i_p, j} \tag{10.25}$$

在候选模型中挑选出误差最小者就是我们的最佳模型。而且，利用运算后的模糊数据时间数列模型来作预测。若模糊集合的结果能满足需求，则预测模型的建立工作就完成了。但是，在许多情况下，所要求的结果往往是数值数据，因而经常需要把模糊集合转换成数值数据，以便利用。在模糊数值转换的过程中，可以遵循下列原则：

(1) 若输出的隶属度只有一个最大值，则我们选择输出最大隶属度，所在区间的中点为输出值。

(2) 若输出的隶属度有两个或更多个连续的最大隶属度，则取最大隶属度，所在区间的典型值平均为输出值。

(3) 其他情况则以标准化后的隶属度为加权数，对各组典型值作平均为其输出值。也就是

$$\text{输出值} = S_1 \times M_1 + \cdots + S_r \times M_r \tag{10.26}$$

其中，S_i 表示标准化后的隶属度，M_i 表示典型值，$i = 1, 2, \cdots, r$。

现在，我们给出建立模糊时间数列模型的算法规则和流程图，如图 10.4 所示。

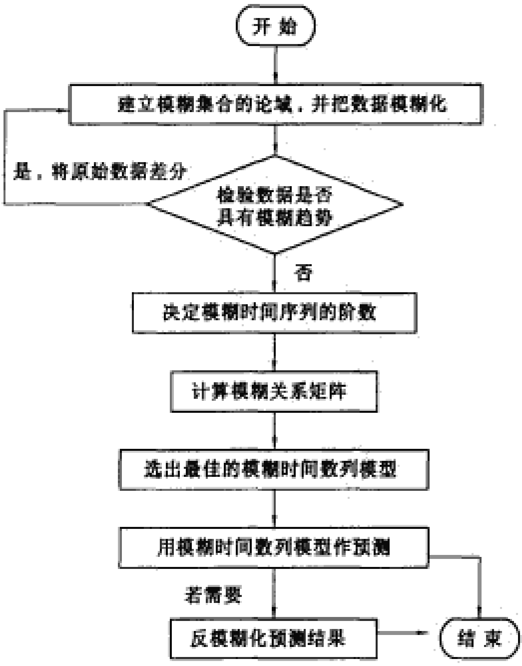


图 10.4　建立模糊时间数列模型的算法关系流程图

建立模糊时间数列模型的算法规则

(1) 建立模糊集合的论域,并把数据模糊化。

(2) 利用检验法则,检验是否具有模糊趋势。若有则将原始数据差分,再回到步骤(1)。否则,继续下一步骤。

(3) 决定模糊时间数列的阶数。

(4) 计算模糊关系矩阵。

(5) 选出最佳的模糊时间数列模型。

(6) 利用运算后的模糊时间数列模型来作预测。

(7) 反模糊化(Defuzzify) 预测结果。

下面,我们给出模糊自回归时间数列阶数的认定算法规则和流程图,如图10.5 所示。

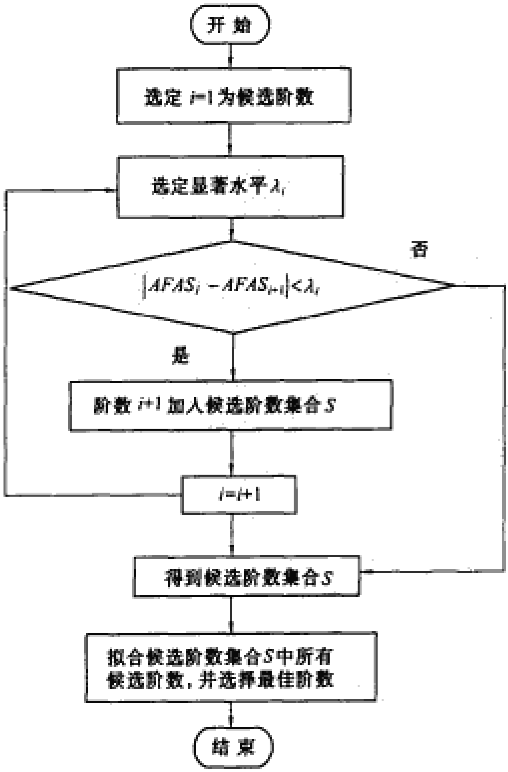


图 10.5 模糊自回归时间数列阶数的认定算法规则和流程图

模糊自回归时间数列阶数判定的算法规则

(1) 设阶数 $i = 1$ 为备选阶数。

(2) 选定显著水平 λ_i。

(3) 若 $|AFAS_i - AFAS_{i+1}| \leqslant \lambda_i$，则阶数 $i+1$ 加入备选阶数集合 S。否则到步骤(4)。

(4) 令 $i = i + 1$，回到步骤(2)。

(5) 得到备选阶数集合 S。

(6) 拟合备选阶数集合 S 所有备选阶数，并选择最佳阶数。

10.5　景气对策信号实例

景气对策信号是中国台湾经济政策最重要的参考指标。因此，此信号是有关单位均希望能预知，以便能事先制定重大经济政策，或企业界作更多的产业投资计划。在 1968 年，获得景气对策信号分数，这些信号分数是经济建设委员会使用趸售物价指数变动率等 12 项统计项目，经过评比所得的综合分数，以 7 个灯号表示。而现在经过 3 次修改，改用货币供给 M_{1B}(日平均) 等 9 项统计项目，用 5 个灯号表示此综合分数。但是，以这些景气指标来作评分，往往需要等到这些统计项目每月指标出来之后，才能作综合分数评判，每次都要等到下月或下下月才能公布当月景气对策信号分数，亦即会落后 1 到 2 期，从而无法预测更多个月的景气对策信号分数。

因此，我们以模糊时间数列分析过程，利用经济建设委员会以往所得的景气对策信号分数，建立一个模糊数据时间数列模型，并预测未来更多期的景气对策信号分数，以便于政府或企业能更早掌握社会的经济情况。现在，就逐步建立模糊数据时间数列，考虑如下：

(1) 由经济建设委员会公布的景气对策信号分数，收集 1992 年到 1996 年，共 60 笔的景气对策信号分数，来建立模糊时间数列模型，并且以此模型预测未来的景气对策信号(分数)。首先，由数据可以绘制其走势图，如图 10.6 所示。

对于景气对策信号，是以下列 5 种灯号来表示其状况。

令 $F_5(t)$ = “红灯”，表示过热，其景气对策信号分数落在 38 以上，以 5 分计算，定为类别 5，以 $I_5 = (38,46]$ 记之；

令 $F_4(t)$ = “黄红灯”，表示尚稳，其景气对策信号分数落在 32 到 38，以 4 分计算，定为类别 4，以 $I_4 = (32,38]$ 记之；

令 $F_3(t)$ = “绿灯”，表示稳定，其景气对策信号分数落在 23 到 32，以 3 分计

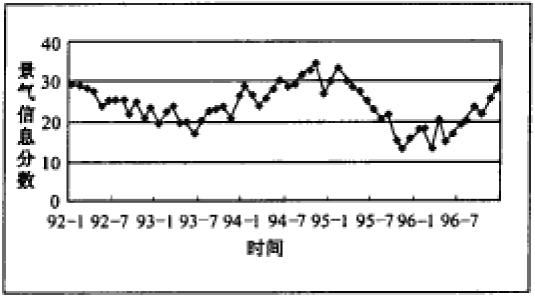


图 10.6 1992 年至 1996 年中国台湾经济景气对策信号分数走势图

算,定为类别 3,以 $I_3 = (23,32]$ 记之;

令 $F_2(t)$ = "黄蓝灯",表示趋衰,其景气对策信号分数落在 17 到 23,以 2 分计算,定为类别 2,以 $I_2 = (17,23]$ 记之;

令 $F_1(t)$ = "蓝灯",表示衰退,其景气对策信号分数落在 17 以下,以 1 分计算定为类别 1,以 $I_1 = (0,17]$ 记之。

其中每一个模糊集的元素为 I_1, I_2, I_3, I_4 以及 I_5 与其相对应的隶属度所组成。由此可知,景气对策信号其实是一组语言的模糊时间数列。因此,使用模糊理论来拟合其模型,应比传统的时间数列方法更为适当。而在我们的观察中,景气对策信号的变化情况,为向上成长,向下衰退,或者维持不变。因而,我们可以假定语言变量集合为

$$F_1(t) = 1/I_1 + 0.5/I_2 + 0/I_3 + 0/I_4 + 0/I_5$$
$$F_2(t) = 0.5/I_1 + 1/I_2 + 0.5/I_3 + 0/I_4 + 0/I_5$$
$$F_3(t) = 0/I_1 + 0.5/I_2 + 1/I_3 + 0.5/I_4 + 0/I_5$$
$$F_4(t) = 0/I_1 + 0/I_2 + 0.5/I_3 + 1/I_4 + 0.5/I_5$$
$$F_5(t) = 0/I_1 + 0/I_2 + 0/I_3 + 0.5/I_4 + 1/I_5$$

其中隶属度的给定,是典型值所在位置给予 1,两旁的隶属度为 0.5。为了方便起见,集合中的元素只使用相对应的隶属度来表示,比如

$$F_1(t) = (1,0.5,0,0,0)$$
$$F_2(t) = (0.5,1,0.5,0,0)$$
$$F_3(t) = (0,0.5,1,0.5,0)$$
$$F_4(t) = (0,0,0.5,1,0.5)$$
$$F_5(t) = (0,0,0,0.5,1)$$

由上面设定可知,原始 60 笔景气对策信号数据其分类区间的走势图,如图 10.7 所示。

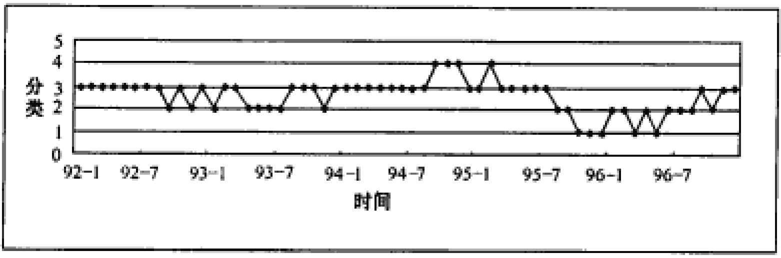


图 10.7　1992 年至 1996 年中国台湾景气对策信号最大隶属度所在分类走势图

(2) 现在，我们要检验此景气对策信号是否具有显著的模糊趋势。根据10.4节的模糊趋势检验法则，其检验过程为：

① 统计假设检验 H_0：没有模糊向下趋势；H_1：有模糊向下趋势；

② 设定显著水平 0.3，检验统计量 $P(F(t-1) >_F F(t)) = 0.5$；

③ 应用单侧检验，在显著水平为 0.3 下，$P(F(t-1) >_F F(t)) < 0.7$；

④ 结论：不拒绝 H_0，亦没有显著的模糊向下趋势。

(3) 因为数据的边界均已经给定，且分为 5 类，因此，我们利用值域均分法的步骤(3)，可得落后 1 至 24 期调整后的模糊自相似度，如图 10.8 所示。

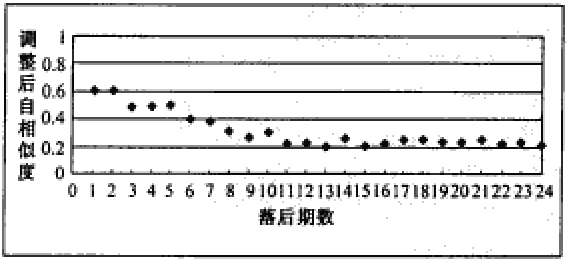


图 10.8　调整后的模糊自相似度

由图 10.8 可以看出，模糊时间数列 $\{F_t\}$ 与 $\{F_{t-1}\}$，$\{F_{t-2}\}$ 有显著的相似。因此，令 $\lambda_1 = 0.05$，$\lambda_2 = 0.0125$，则由阶数的判定法则知，阶数可能为 1 或 2。

(4) 我们利用标准的隶属函数，将景气对策信号分数数据模糊化，亦即在式(10.26) 的每个集合 $F_i(t)(i = 1,2,\cdots,5)$，针对每月的数据找出对应的隶属度。其相对应的隶属函数，如图 10.9 所示。

由图 10.9，我们利用标准的隶属函数，把景气对策信号分数数据模糊化，可得其相对应的隶属度集合，其集合的一部分如表 10.4 所示。

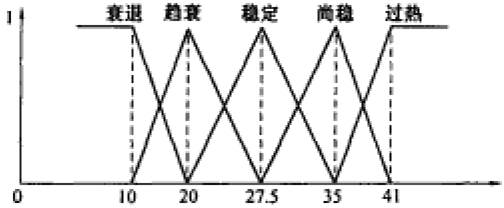


图 10.9　景气对策信号模糊化的隶属函数

表 10.4　每月景气信号的隶属函数

年／月	$F_1(t)$	$F_2(t)$	$F_3(t)$	$F_4(t)$	$F_5(t)$
1992/1	0	0	0.8	0.2	0
1992/2	0	0	0.8	0.2	0
1992/3	0	0	0.93	0.07	0
⋮	⋮	⋮	⋮	⋮	⋮
1996/10	0	0.73	0.27	0	0
1996/11	0	0.2	0.8	0	0
1996/12	0	0	0.93	0.07	0

由表 10.4 可得每月景气信号的历史经验。假定每月景气信号的最大隶属度是在 $F_i(t)(i=1,2,\cdots,5)$，则认为这月的景气信号为 $F_i(t)$。例如，以 1992 年 3 月为例，最大的隶属度是在 F_3(1992/3)，则可认为 1992 年 3 月的景气信号为 F_3(1992/3)，或称之为是属于"景气信号稳定" 的月份。

通过许多过去的模糊数据，理论上可以找出模糊关系矩阵。由阶数的认定告诉我们，当期的数据可能受前一期或前两期的影响，亦即为"若 $F(t-1)$ 是 A，则 $F(t)$ 是 B" 或"若 $F(t-2)$ 是 A_1 或 $F(t-1)$ 是 A_2，则 $F(t)$ 是 B" 的概念。有鉴于此，可以分别就模糊集合中探讨连续两个月或三个月景气信号的模糊关系。假定"第 $t-1$ 月的景气信号为 $F_m(t-1)$，则第 t 月的景气信号为 $F_n(t)$"。它们之间的模糊关系可以 $F_m(t-1)\rightarrow F_n(t)$ 表示。或假定"第 $t-2$ 月的景气信号为 $F_{m_1}(t-2)$ 或第 $t-1$ 月的景气信号为 $F_{m_2}(t-1)$，则第 t 月的景气信号为 $F_n(t)$"。它们之间的模糊关系可以"$F_{m_1}(t-2)$ 或 $F_{m_2}(t-1)\rightarrow F_n(t)$" 表示，其中 $m,n,m_1,m_2\in\{1,2,3,4,5\}$。

因此，可以获得景气信号的模糊关系分别如下（在此，我们把时间指标省略了）：

(a) 若 $F_m(t-1)$ 是 A，则 $F_n(t)$ 是 B

$$F_3 \to F_3, F_3 \to F_2, F_2 \to F_3, F_2 \to F_2, F_2 \to F_1$$
$$F_1 \to F_2, F_3 \to F_4, F_4 \to F_4, F_4 \to F_3, F_1 \to F_1$$

(b) 若 $F_{m_1}(t-2)$ 是 A_1 或 $F_{m_2}(t-1)$ 是 A_2，则 $F_n(t)$ 是 B

$F_3 \cup F_3 \to F_3, F_3 \cup F_3 \to F_2, F_2 \cup F_3 \to F_2, F_3 \cup F_2 \to F_3,$
$F_3 \cup F_2 \to F_2, F_2 \cup F_2 \to F_2, F_2 \cup F_2 \to F_3, F_2 \cup F_3 \to F_3,$
$F_3 \cup F_3 \to F_4, F_3 \cup F_4 \to F_4, F_4 \cup F_4 \to F_4, F_4 \cup F_4 \to F_3,$
$F_4 \cup F_3 \to F_3, F_3 \cup F_4 \to F_3, F_2 \cup F_2 \to F_1, F_2 \cup F_1 \to F_1,$
$F_1 \cup F_1 \to F_1, F_1 \cup F_1 \to F_2, F_1 \cup F_2 \to F_2, F_2 \cup F_1 \to F_2,$
$F_1 \cup F_2 \to F_1, F_2 \cup F_2 \to F_2.$

若我们所定义的模型为一时常值的模糊时间数列。

(a′) 若 $F_m(t-1)$ 是 A，则 $F_n(t)$ 是 B^T，令

$R_1 = F_3 \times F_3^T, R_2 = F_3 \times F_2^T, R_3 = F_2 \times F_3^T, R_4 = F_2 \times F_2^T$
$R_5 = F_2 \times F_1^T, R_6 = F_1 \times F_2^T, R_7 = F_3 \times F_4^T, R_8 = F_4 \times F_4^T$
$R_9 = F_4 \times F_3^T, R_{10} = F_1 \times F_1^T$

(b′) 若 $F_{m_1}(t-2)$ 是 A_1 或 $F_{m_2}(t-1)$ 是 A_2，则 $F_n(t)$ 是 B^T，令

$R_1 = F_3 \times F_3^T, R_2 = F_3 \times F_2^T, R_3 = F_2 \times F_2^T \cup F_3 \times F_2^T$
$R_4 = F_3 \times F_3^T \cup F_2 \times F_3^T, R_5 = F_2 \times F_2^T$
$R_6 = F_2 \times F_3^T \cup F_2 \times F_2^T, R_7 = F_3 \times F_4^T$
$R_8 = F_4 \times F_4^T, R_9 = F_4 \times F_3^T$
$R_{10} = F_4 \times F_3^T \cup F_3 \times F_3^T, R_{11} = F_2 \times F_1^T$
$R_{12} = F_2 \times F_1^T \cup F_1 \times F_1^T, R_{13} = F_1 \times F_1^T$
$R_{14} = F_1 \times F_2^T, R_{15} = F_1 \times F_2^T \cup F_2 \times F_2^T$
$R_{16} = F_1 \times F_1^T \cup F_2 \times F_1^T, R_{17} = F_2 \times F_2^T$

其中，× 表示取最小运算，T 表示转置。然后，可以得到

$$\boldsymbol{R} = \sup_{1 \leq i \leq 10} R_i$$

或

$$\boldsymbol{R} = \sup_{1 \leq i \leq 17} R_i$$

一旦利用式(10.26)，经过计算之后可得 1 阶模糊关系矩阵 $\boldsymbol{R}$ 为

$$\boldsymbol{R} = \begin{pmatrix} 1 & 1 & 0.5 & 0.5 & 0 \\ 1 & 1 & 1 & 0.5 & 0.5 \\ 0.5 & 1 & 1 & 1 & 0.5 \\ 0.5 & 0.5 & 1 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}$$

或 2 阶模糊关系矩阵 $\boldsymbol{R}$ 为

$$\boldsymbol{R}=\begin{pmatrix}1 & 1 & 0.5 & 0.5 & 0\\ 1 & 1 & 1 & 0.5 & 0.5\\ 0.5 & 1 & 1 & 1 & 0.5\\ 0.5 & 0.5 & 1 & 1 & 0.5\\ 0 & 0.5 & 0.5 & 0.5 & 0.5\end{pmatrix}$$

(5) 预测模型可能为

$$F(t)=F(t-1)\circ R \tag{10.27}$$

或

$$F(t)=F(t-1)\circ F(t-2)\circ R \tag{10.28}$$

其中，$F(t-2)$，$F(t-1)$，$F(t)$ 分别表示模糊集合中第 $t-2$ 月，第 $t-1$ 月，第 t 月的景气对策信号。最后，计算出模型输出的隶属度，并将隶属度予以标准化。处理的结果如表 10.5 及表 10.6 所示。

表 10.5　$FAR(1)$ 的景气对策信号预测的标准化隶属度与拟合灯号(值)

年／月	标准化隶属度	拟合灯号(值)
1992/2	0.14,0.24,0.24,0.24,0.14	绿灯(27.5)
1992/3	0.14,0.24,0.24,0.24,0.14	绿灯(27.5)
1992/4	0.13,0.24,0.26,0.24,0.13	绿灯(27.5)
⋮	⋮	⋮
1996/10	0.2,0.2,0.2,0.2,0.2	绿灯(27.5)
1996/11	0.22,0.24,0.22,0.16,0.16	黄蓝灯(20.5)
1996/12	0.14,0.24,0.24,0.24,0.14	绿灯(27.5)

表 10.6　$FAR(2)$ 的景气信号预测的标准化隶属度与拟合灯号(值)

年／月	标准化隶属度	拟合灯号(值)
1992/3	0.125,0.25,0.25,0.25,0.125	绿灯(27.5)
1992/4	0.125,0.25,0.25,0.25,0.125	绿灯(27.5)
1992/5	0.22,0.22,0.22,0.19,0.16	黄蓝灯(20)
⋮	⋮	⋮
1996/10	0.16,0.2,0.24,0.24,0.16	绿灯(31.25)
1996/11	0.19,0.19,0.23,0.23,0.16	绿灯(31.25)
1996/12	0.19,0.19,0.23,0.23,0.16	绿灯(31.25)

(6) 拟合结果。

景气信号分数 $FAR(1)$ 模型的拟合结果，如图 10.10 所示。

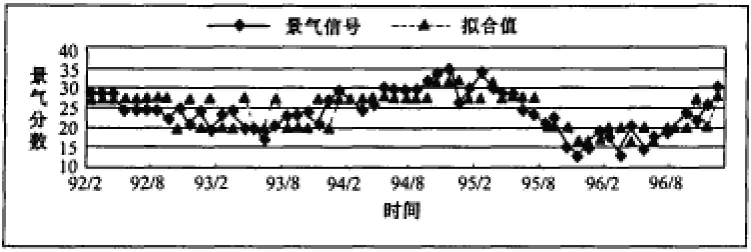


图 10.10　景气信号 $FAR(1)$ 拟合值与真实值曲线

景气信号分数 $FAR(2)$ 模型拟合的结果，如图 10.11 所示。

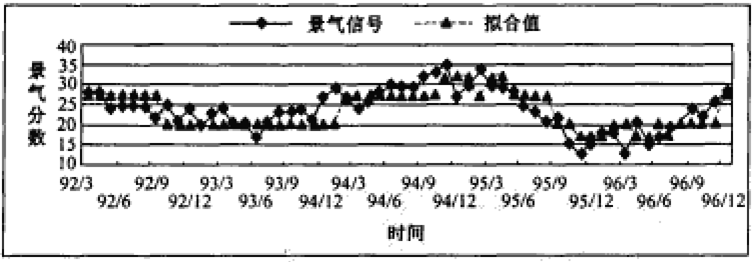


图 10.11　景气信号 $FAR(2)$ 拟合值与真实值曲线

一般地说，我们以均方误差(Mean Squared Error，MSE)、平均绝对误差(MAE)以及平均绝对百分误差率(MAPE)来做拟合比较标准。其公式如下

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\hat{x}_i - x_i)^2 \tag{10.29}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|\hat{x}_i - x_i| \tag{10.30}$$

$$MAPE = \frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{\hat{x}_i - x_i}{x_i}\right| \tag{10.31}$$

由表 10.7 可得，$FAR(2)$ 的 $MSE = 12.6$ 及 $MAE = 12.8\%$，均比 $FAR(1)$ 的 $MSE = 10.7$ 及 $MAE = 10.4\%$ 要小一些。因此，我们选择最佳模型为 $FAR(2)$。

(7) 预测模型式(10.27)及式(10.28)的输出结果为模糊集合。若模糊集合的结果能满足需求，则模糊时间数列模型的建立工作便告完成了。但是，在许多情况下，所要求的结果往往是数值数据，所以往往也需要将模糊集合转换成数值数据，以便于利用。根据 10.5 节，可以获得拟合的输出值，如表 10.5 及表 10.6 所示。另外，未来一年的景气分数预测值，如表 10.8 所示。

表 10.7　景气分数的 *FAR*(1) 与 *FAR*(2) 拟合误差比较

	FAR(1)	*FAR*(2)
MSE	10.7	12.6
MAE	3.15	2.91
MAPE	10.4%	12.8%

表 10.8　1997 年景气对策信号的真实值与预测值

年／月	真实值	*FAR*(1) 预测值	*FAR*(2) 预测值
1997/1	绿灯(26)	绿灯(27.5)	绿灯(27.5)
1997/2	绿灯(24)	绿灯(27.5)	绿灯(27.5)
1997/3	绿灯(31)	绿灯(27.5)	绿灯(27.5)
1997/4	绿灯(24)	绿灯(27.5)	绿灯(27.5)
1997/5	绿灯(24)	绿灯(27.5)	绿灯(27.5)
1997/6	绿灯(28)	绿灯(27.5)	绿灯(27.5)
1997/7	绿灯(30)	绿灯(27.5)	绿灯(27.5)
1997/8	绿灯(26)	绿灯(27.5)	绿灯(27.5)
1997/9	绿灯(31)	绿灯(27.5)	绿灯(27.5)
1997/10	绿灯(26)	绿灯(27.5)	绿灯(27.5)
1997/11	绿灯(28)	绿灯(27.5)	绿灯(27.5)
1997/12	绿灯(28)	绿灯(27.5)	绿灯(27.5)

由表 10.8 可得，*FAR*(1) 与 *FAR*(2) 的预测值均相同，且都能准确地预测灯号。而其预测值之所以均为 27.5，是因为模糊时间数列为一区间预测。另外，每一应用有其适合的反模糊化方法，参看朗克尔(Runkler,1997)。若能找出更适合的反模糊化方法，则将能更准确地预测结果。

10.6　结　论

在时间数列分析过程中，数据的模糊性和模糊相关性的问题，可以说是建立传统时间数列所不易克服的困难。而若结合较稳健性与符合实际状况的模糊统计分析，应可避免这些隐藏的问题。事实上，在人类思维模型中，计量过程与标准皆充满着模糊性。一方面，这似乎是单纯化的结果；另一方面，过分单纯化却又阻碍计量方法与标准的客观性。

本章给出一些定义，比如定义 10.4 模糊高于及定义 10.5 模糊趋势，以此建立模糊趋势检验法则；定义 10.8 提出模糊自相似度的概念；定义 10.9 调整后模糊自相似度的概念，以此建立模糊阶数的认定法则等，这些均是本章提出的重要结果；定义 10.10 则明确说明模糊自回归时间数列。而其中最重要的结论是，我们建立了模糊自回归时间数列，确立模糊趋势型时间数列的程序。同时，对于模糊趋势型时间数列，提出有效的判断法则。另外，对于模糊趋势型时间数列进行定阶亦确立了良好的算法规则。

实际上，目前仍有一些需要继续进一步探讨及研究的问题与方向，例如：

当系数有正亦有负的情况，或全为负的情况，其模糊自回归阶数如何认定？

阶数为 2 以上的情况，模糊参数的推导相当繁杂，且其阶数如何认定？

如何发展出比较简便的程序？

对于模糊自回归时间数列及模糊季节性时间数列同时存在的阶数如何进行定阶？

对于非线性的时间数列，其相对应的模糊时间数列的阶数如何进行定阶？

对于多变量的模糊时间数列，其相对应的模糊时间数列的阶数如何进行定阶？模型如何建立？

第 11 章　问题及展望

在过去的二十年里，特别是近十年来，出版的大量的论文致力于扩展已有的概率论及统计学(包括数理统计学)。这些研究及探索性工作的一个共同特性是对经典理论进行“软化”。某些软化方法使用了诸如模糊集合理论、可能性理论，还有一些其他方法，比如 Dempster-Shafer 理论(又称证据理论)，而这些方法都起源于经典概率论。

在一些广泛的应用领域中，对概率论进行软化的推广需求显而易见。例如，在数值分析与数据挖掘方面，一旦把模糊集合与概率论加以整合，就能导致更好地既捕获到基本数据的内在不确定性，又捕获到基本数据的模糊性，这类探索工作日益增多。

例如，在科学和工程的研究及应用中，对与复杂系统有关的真实不确定性的分析及建模的企图和研究需要广义不确定性理论、广义随机元、广义概率、模糊概率及统计等等。

不确定性对统计推断的影响

统计推理是在部分信息下进行推理的一种方法，因而也是在不确定性下实施的推理。

最近，意大利学者 R. Coppi 教授在 2002 年“数据挖掘的理论框架”一文中提出统计学的信息范式(Information Paradigm)，认为统计推理的思想是从信息中提取信息，并适当管理相关的不确定之一系列逻辑与数学程序。更详细地讲，可以把统计关注的信息分成两大主要类型。一方面，我们具有构成数据的经验信息(Empirical Information, S_E)，也就是汇总的现象数据 —— 与事例有关的所捕获到的那些汇总的现象数据；另一方面，我们拥有理论信息(Theoretical Information, S_T)，这既包括支撑进行数据分析的理论假设，也包括详尽阐述数据的(算法) 程序。统计推理作用于序对(S_E, S_T)，以此获得“额外信息”，该额外信息的基本来源根植于汇总数据，而这会受到各种各样类型不确定性的影响。因此，通过统计推理所提取的额外信息自身也会受到不确定性的影响。而后者不确定性是相关的各种不确定性传播的结果，尽管统计推理的特定系统在上述

每一种情况下都要进行修正。

根据这种“信息范式”，极为重要的是如何去设定在当今广泛统计学观点中起着重大作用的各类不确定性。这既有助于我们认清、梳理各种理论上及方法上的进展与业已取得的成果，又有助于遴选出将来需要进一步建立更系统的(包括统计方法)方法论的真知灼见。

基于上述的信息范式，确有必要研究各种各样的不确定性类型，因为它们会影响到构成统计推理过程不同阶段的信息体(既有经验上的，又有理论上的)。对不确定性类型进行分析，并加以设定的一种合乎逻辑的方式如下：

(1) 与可观测数据和“可能数据”领域之间联系的不确定性；例如，在传统推断方法中，这通过概率抽样模型来加以管理。

(2) 测度经验现象中所产生的不准确性(不确定性)；例如，这会产生不准确的统计数据，仅仅通过忽略其内在的不准确性而处理成“明确”数据。

(3) 以语言术语描述现实世界的相关模糊性(不确定性)；例如，分类数据或有序数据最经常包括这类不确定性。然而，注意到，某种理论上推理成分可以“模糊方式”进行定义，比如，统计关系或者统计假设的表述。

(4)(部分或者全部)忽略特定观测事例的现象值(比如，缺失数据)，或者对给定理论假设的部分或者全部忽略；例如，当对随机量值采用特定概率模型时，在传统统计方法论中，一些特定方法被设计成用于处理这种不确定性的类型，诸如稳健统计量，或者基于对统计模型的不确定特征引入先验概率分布的贝叶斯分析。

(5) 源于描述物质世界时所使用的“颗粒化”术语的不准确性；这与语言变量的一般概念有关，语言变量的“颗粒性”(Granules)是指通过难区分性、相似性、近似性以及功能性所引申出的一些字词。这个概念比定性变量更为一般，它们在统计学中得到了应用。实际上可以把它归属于包括对模型的定义与使用、理论假设等认知过程，而这也是所谓的统计推理中的理论信息。

由上述统计学与不确定性联系看，显然，在给定的不确定性类型与相关的信息成分(既有经验上的，又有理论上的)之间存在着直接的对应。在传统的推断统计学背景下，所列出的一些不确定性借助于适当的修正过程而得到控制。目前，统计学家大体上仍然沿着这种路线在研究及探索。

然而，不确定性理论的丰硕成果正在开启着统计学家直接面对统计学中不确定性而进行探索的一种新方式。

模糊集合理论对统计学的渗透

模糊集合理论的创立及其在逻辑学、数学、工程学等领域的不断发展和传

播为统计学方法论提供了富有成果的新思想、新工具。从20世纪60年代中期开始，日益增多的研究成果扩充了统计推理的范围，进而包括模糊信息及模糊不确定性。

模糊数学研究方法及思想对统计分析及建模产生了深远而广泛的影响，如同前言所述，对不确定性或不准确性的关注已经渗透到统计理论分析的信息成分中，比如参数或模型的其他特性上。

几个最为成功的事例是模糊聚类(Fuzzy Clustering)，比如贝兹德克(Bezdek，1981)和模糊回归分析(Fuzzy Regression Analysis)，比如田中等(Tanaka，et al，1982)。

模糊聚类

设我们考察取值 n 个统计单元的一些可观测值，通过 p 个数量变量 $x'_i = (x_{i1},\cdots,x_{ip})$ 来表示，$i = 1,2,\cdots,n$。在此经验数据基础上，希望确定 K 个群或类(Clusters)，比如说 R^p 中的 $C_1,\cdots,C_k,\cdots,C_K$ 来揭示出每一个类"模糊拓扑(Fuzzy Typology)"所刻画的一些单元的"象征性"。通常，$C_k(k = 1,2,\cdots,K)$ 原型可以用 R^p 中的向量 c_k 表示。

在以往的聚类分析中，"明确"聚类目标是对每一个类都要顾及其经典特征函数 u_{ik} 在[0,1]中的取值，也就是对于群来说，允许我们将单元 i 指派到类 $k(u_{ik} = 1)$ 或不知派到类 $k(u_{ik} = 0)$，与此同时，其目的是为我们提供原型的估计值，绝大多数情况下，常常是给出这些类的形心(Centroids，又称质心、重心)。

上述聚类建模方式可以适当地通过模糊方法加以推广。因此，如果强调相关的不准确性或模糊性，那么在更现实研究中就要把模糊思想融入到分类方法中。这可通过定义在 n 个单元集合上的 $u_{ik}(k = 1,2,\cdots,K)$ 成为相对应的模糊类的隶属函数来直接完成。在前面框架下，一种最为流行的方法是由贝兹德克(1981)所建立的模糊 K 均值法。该方法是建立在对下述目标函数

$$F(U,C \mid X,m,K) = d^2(x_i,c_k)$$

求最小值基础上的一种非递阶技术(Nonhierarohical Technique)，其中 U 表示隶属函数矩阵，$C = (c_1,\cdots,c_K)$ 表示类形心的矩阵，X 表示数据矩阵，而 d 表示 R^p 中的适当距离函数。注意到，上式依赖于两个应预先设定的参数，也就是模糊系数 $m(m \geq 1)$，它表示所研究原型的模糊性程度，还有模糊类的个数 K，它表示所要探讨的模糊聚类之分类数。

迄今为止，模糊聚类分析方法的成果及影响是至今所研究模糊数据统计分析与建模中最为盛行而成功的。实际上，同样可以考察"模糊样板(Fuzzy

Prototypes)" 的情况，这就产生了模糊向量刻画模糊类。这适用于那些可观测数据本身是模糊数的情况。例如，可参看 D'Urso 和 Giordani(2006)。

模糊回归分析

在传统回归分析中，观测值与估计值之间的离差都被假定成起因于随机误差。然而，实际上，它们常常起因于方程结构的模糊性或者不准确的观测值。因而，这类回归模型中的不确定性就是模糊性及非随机性。目前，关于模糊线性回归(Fuzzy Linear Regression, FLR) 研究文献大致分成两种方法：一种是基于线性规划方法(Linear Programming – based (LP) Methods)，另一种是模糊最小二乘法(Fuzzy Least Squares (FLS) Methods)。

第一种方法是由田中等人(1982) 提出的，FLR 参数模型是

$$Y = A_0 + A_1X_1 + \cdots + A_pX_p$$

其中 $A_i = (a_{ic}, a_{iw})$, $i = 1, \cdots, p$, $Y = (y_c, y_w)$ 均可通过对模糊数据组合的总模糊性求最小值而得到估计，使得每个数据点必须位于响应变量的估计值之内。这可以想象成为一个 LP 问题。并且通过利用"单纯形法" 来求解。一些软件包，诸如 SAS, LP88 以及 LINDO 都可用于求解 LP 问题。任何标准的电子表软件，比如微软的 Excel 也可以用于以手工方式加以求解。

然而，田中方法的缺陷是，它没有建立在坚实的统计学原理之上，而且另外一个缺点是，当数据点增多时，约束数目会成比例增长，因而导致计算上的困难。例如，参看 Chang 和 Ayyub(2001)，D'Urso(2003)。

第二种方法是建立在模糊最小二乘法(FLS) 之上，顾名思义，它是建立在模糊数空间中新定义距离基础之上对最小二乘法的模糊推广，最初是由戴蒙德(Diamond, 1988) 提出的。然而，这种方法的缺陷是，即使观测响应的幅度大致为常值或者减少时，估计响应的波动幅度仍会随着解释变量数量的增多而增大。为了克服这一缺点，Kao 和 Chyu(2002) 通过 FLS 方法提出了拟合 FLP 模型的两阶段方法，并且证明，它优于戴蒙德方法。Yang 和 Liu(2003) 对 FLP 模型中存在异常值时，提出一种稳健算法。

与模糊线性模型相比，至今论述模糊非线性模型(Fuzzy Nonlinear Models, FNMs) 的文献极少。为了拟合某种特殊参数的 FNMs，Buckley 和 Feuring(2000) 提出"进化算法求解"(Evolutionary Algorithm Solutions)。特别地，对于给定模糊数据，从模糊函数库(包括线性的、多项式的、指数的以及对数函数) 寻找并探索最佳拟合数据的函数，显然，这并不是对所有参数拟合数据来说都是令人满意的。因此，这仍是一个未来需要进一步探索及发展的专题。

Hong 和 Hwang(2003) 引进支持向量机制(Support Vector Machines, SVM)来拟合非参数的FNMs。实际上,SVM最初是由AT&T贝尔实验室的 Vapnik 及其合作者所发展起来的(Vapnik,2000),它是建立在结构风险最小化思想基础上的。Suyken 等人(2002) 对 SVM 给出一个极好的综述。然而,为了把 SVM 用于现实数据上,还需要作出大量的推广精细调整的方法研究工作。

模糊时间序列

在预测应用中,时间序列模型是最普遍使用的工具之一。然而,如果给定数据是语言变量,那么这种统计方法将失效。为了研究此类问题,模糊时间序列模型被发展起来,并用于实践中。Song 和 Chisson 是研究词类问题的先驱,他们在1993 年曾提出模糊时间序列模型。由于在某些类型预测问题中,它表现出较好的效果,因此模糊时间序列模型受到研究者的青睐。受到 Song 和 Chisson 方法的启发,Lee 曾提出对具有三角形模糊数的历史数据进行处理的方法,同时使用 Song 和 Chisson 的时间变式模糊时间序列模型来预测大学注册问题。此外,Hwang 应用历史数据的变差代替 Song 和 Chisson 的方法中的最初输入,并获得了合情合理的结果。

实际上,如同 10.6 节所述,目前模糊时间序列模型建模中仍存在许多问题,比如对于模糊自回归时间数列及模糊季节性时间数列同时存在的阶数如何进行定阶等。建立模糊时间序列模型及进行预测的主要问题在于,对模糊关系的建立与计算,还有确立从所得出的模糊输出中获得明确输出的去模糊化的合适方法。

总之,模糊数据统计学正处于蓬勃发展中,日益成长壮大,涌现出一些探索模糊性与经典统计方法融合的新思想、新方法,这是一个值得人们开拓耕耘、不断努力探索的新交叉性领域。

附　　录

附录 1　偏大与偏小型隶属函数表

1. 偏小型隶属函数

(1) 降半矩阵分布

$$\mu(x)=\begin{cases}1 & x\leqslant a\\ 0 & x>a\end{cases}$$

参看图 1。

(2) 降半 Γ 分布

$$\mu(x)=\begin{cases}1 & x\leqslant a\\ e^{-k(x-a)} & x>a,k>0\end{cases}$$

参看图 2。

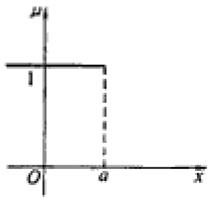


图 1　降半矩阵分布

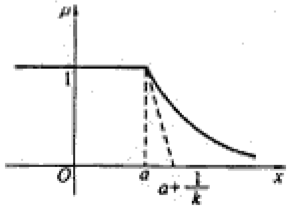


图 2　降半 Γ 分布

(3) 降半正态分布

$$\mu(x)=\begin{cases}1 & x\leqslant a\\ e^{-k(x-a)^2} & x>a,k>0\end{cases}$$

参看图 3。

(4) 降半柯西分布

$$\mu(x)=\begin{cases}1 & x\leqslant a\\ \dfrac{1}{1+\alpha(x-a)^{\beta}} & x>a,\alpha>0,\beta>0\end{cases}$$

参看图 4。

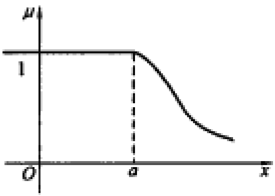


图 3　降半正态分布

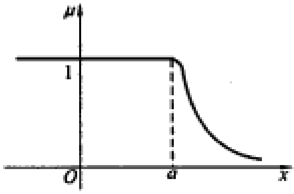


图 4　降半柯西分布

(5) 降半梯形分布

$$\mu(x)=\begin{cases}1 & x\leqslant a_1\\ \dfrac{a_2-x}{a_2-a_1} & a_1<x\leqslant a_2\\ 0 & a_2<x\end{cases}$$

参看图 5。

(6) 降岭形分布

$$\mu(x)=\begin{cases}1 & x\leqslant a_1\\ \dfrac{1}{2}-\dfrac{1}{2}\sin\dfrac{\pi}{a_2-a_1}\left[x-\dfrac{a_1+a_2}{2}\right] & a_1<x\leqslant a_2\\ 0 & a_2<x\end{cases}$$

参看图 6。

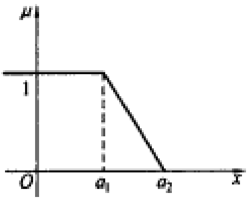


图 5　降半梯形分布

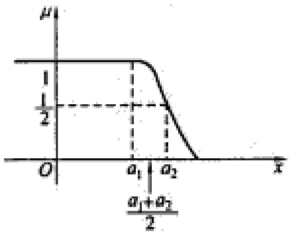


图 6　降岭形分布

2. 偏大型隶属函数

(1) 升半矩阵分布

$$\mu(x)=\begin{cases}0 & x\leqslant a\\ 1 & x>a\end{cases}$$

参看图 7。

(2) 升半 Γ 分布

$$\mu(x)=\begin{cases}0 & x\leqslant a\\ 1-e^{-k(x-a)} & x>a,k>0\end{cases}$$

参看图 8。

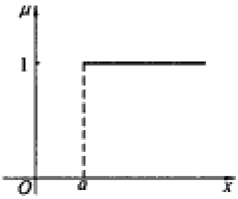


图 7　升半矩阵分布

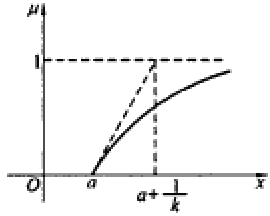


图 8　升半 Γ 分布

(3) 升半正态分布

$$\mu(x)=\begin{cases}0 & x\leqslant a\\ 1-e^{-k(x-a)^2} & x>a,k>0\end{cases}$$

参看图 9。

(4) 升半柯西分布

$$\mu(x)=\begin{cases}0 & x\leqslant a\\ \dfrac{1}{1+\alpha(x-a)^{-\beta}} & x>a,\alpha>0,\beta>0\end{cases}$$

参看图 10。

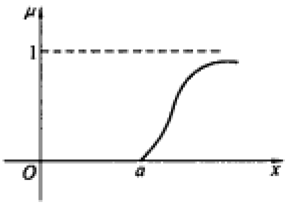


图 9　升半正态分布

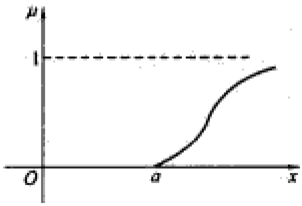


图 10　升半柯西分布

(5) 升半梯形分布

$$\mu(x)=\begin{cases}0 & x\leqslant a_1\\ \dfrac{x-a_1}{a_2-a_1} & a_1<x\leqslant a_2\\ 1 & a_2<x\end{cases}$$

参看图 11。

(6) 升岭形分布

$$\mu(x)=\begin{cases}0 & x\leqslant a_1\\ \dfrac{1}{2}+\dfrac{1}{2}\sin\dfrac{\pi}{a_2-a_1}\left[x-\dfrac{a_1+a_2}{2}\right] & a_1<x\leqslant a_2\\ 1 & a_2<x\end{cases}$$

参看图 12。

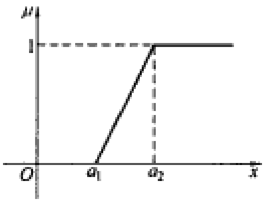


图 11　升半梯形分布

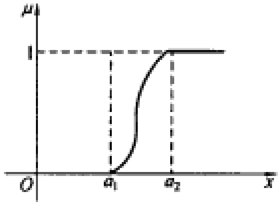


图 12　升岭形分布

附录 2　二项分布累积概率 $P(X \leqslant x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}$

$n = 5$

k	p														
	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.951	0.774	0.590	0.328	0.237	0.168	0.078	0.031	0.010	0.002	0.001	0.000	0.000	0.000	0.000
1	0.999	0.977	0.919	0.737	0.633	0.528	0.337	0.187	0.087	0.031	0.016	0.007	0.000	0.000	0.000
2	1.000	0.999	0.991	0.942	0.896	0.837	0.683	0.500	0.317	0.163	0.104	0.058	0.009	0.001	0.000
3	1.000	1.000	1.000	0.993	0.984	0.969	0.913	0.812	0.663	0.472	0.367	0.263	0.081	0.023	0.001
4	1.000	1.000	1.000	1.000	0.999	0.998	0.990	0.969	0.922	0.832	0.763	0.672	0.410	0.226	0.049

$n = 6$

k	p														
	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.941	0.735	0.531	0.262	0.178	0.118	0.047	0.016	0.004	0.001	0.000	0.000	0.000	0.000	0.000
1	0.999	0.967	0.886	0.655	0.534	0.420	0.233	0.109	0.041	0.011	0.005	0.002	0.000	0.000	0.000
2	1.000	0.998	0.984	0.901	0.831	0.744	0.544	0.344	0.179	0.070	0.038	0.017	0.001	0.000	0.000
3	1.000	1.000	0.999	0.983	0.962	0.930	0.821	0.656	0.456	0.256	0.169	0.099	0.016	0.002	0.000
4	1.000	1.000	1.000	0.998	0.995	0.989	0.959	0.891	0.767	0.580	0.466	0.345	0.114	0.033	0.001
5	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.984	0.953	0.882	0.822	0.738	0.469	0.265	0.059

$n = 7$

k	p														
	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.932	0.698	0.478	0.210	0.133	0.082	0.028	0.008	0.002	0.000	0.000	0.000	0.000	0.000	0.000
1	0.998	0.956	0.850	0.577	0.445	0.329	0.159	0.063	0.019	0.004	0.001	0.000	0.000	0.000	0.000
2	1.000	0.996	0.974	0.852	0.756	0.647	0.420	0.227	0.096	0.029	0.013	0.005	0.000	0.000	0.000
3	1.000	1.000	0.997	0.967	0.929	0.874	0.710	0.500	0.290	0.126	0.071	0.033	0.003	0.000	0.000
4	1.000	1.000	1.000	0.995	0.987	0.971	0.904	0.773	0.580	0.353	0.244	0.148	0.026	0.004	0.000
5	1.000	1.000	1.000	1.000	0.999	0.996	0.981	0.937	0.841	0.671	0.555	0.423	0.150	0.044	0.002
6	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.992	0.972	0.918	0.867	0.790	0.522	0.302	0.068

$n=8$

k	p														
	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.923	0.663	0.430	0.168	0.100	0.058	0.017	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000
1	0.997	0.943	0.813	0.503	0.367	0.255	0.106	0.035	0.009	0.001	0.000	0.000	0.000	0.000	0.000
2	1.000	0.994	0.962	0.797	0.679	0.552	0.315	0.145	0.050	0.011	0.004	0.001	0.000	0.000	0.000
3	1.000	1.000	0.995	0.944	0.886	0.806	0.594	0.363	0.174	0.058	0.027	0.010	0.000	0.000	0.000
4	1.000	1.000	1.000	0.990	0.973	0.942	0.826	0.637	0.406	0.194	0.114	0.056	0.005	0.000	0.000
5	1.000	1.000	1.000	0.999	0.996	0.989	0.950	0.855	0.685	0.448	0.321	0.203	0.038	0.006	0.000
6	1.000	1.000	1.000	1.000	1.000	0.999	0.991	0.965	0.894	0.745	0.633	0.497	0.187	0.057	0.003
7	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.983	0.942	0.900	0.832	0.570	0.337	0.077

$n=9$

k	p														
	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.914	0.630	0.387	0.134	0.075	0.040	0.010	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.997	0.929	0.775	0.436	0.300	0.196	0.071	0.020	0.004	0.000	0.000	0.000	0.000	0.000	0.000
2	1.000	0.992	0.947	0.738	0.601	0.463	0.232	0.090	0.025	0.004	0.001	0.000	0.000	0.000	0.000
3	1.000	0.999	0.992	0.914	0.834	0.730	0.483	0.254	0.099	0.025	0.010	0.003	0.000	0.000	0.000
4	1.000	1.000	0.999	0.980	0.951	0.901	0.733	0.500	0.267	0.099	0.049	0.020	0.001	0.000	0.000
5	1.000	1.000	1.000	0.997	0.990	0.975	0.901	0.746	0.517	0.270	0.166	0.086	0.008	0.001	0.000
6	1.000	1.000	1.000	1.000	0.999	0.996	0.975	0.910	0.768	0.537	0.399	0.262	0.053	0.008	0.000
7	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.980	0.929	0.804	0.700	0.564	0.225	0.071	0.003
8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.990	0.960	0.925	0.866	0.613	0.370	0.086

$n=10$

k	p														
	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.904	0.599	0.349	0.107	0.056	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.996	0.914	0.736	0.376	0.244	0.149	0.046	0.011	0.002	0.000	0.000	0.000	0.000	0.000	0.000
2	1.000	0.988	0.930	0.678	0.526	0.383	0.167	0.055	0.012	0.002	0.000	0.000	0.000	0.000	0.000
3	1.000	0.999	0.987	0.879	0.776	0.650	0.382	0.172	0.055	0.011	0.004	0.001	0.000	0.000	0.000
4	1.000	1.000	0.998	0.967	0.922	0.850	0.633	0.377	0.166	0.047	0.020	0.006	0.000	0.000	0.000
5	1.000	1.000	1.000	0.994	0.980	0.953	0.834	0.623	0.367	0.150	0.078	0.033	0.002	0.000	0.000
6	1.000	1.000	1.000	0.999	0.996	0.989	0.945	0.828	0.618	0.350	0.224	0.121	0.013	0.001	0.000
7	1.000	1.000	1.000	1.000	1.000	0.998	0.988	0.945	0.833	0.617	0.474	0.322	0.070	0.012	0.000
8	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.989	0.954	0.851	0.756	0.624	0.264	0.086	0.004
9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.972	0.944	0.893	0.651	0.401	0.096

$n = 15$

k	p														
	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.860	0.463	0.206	0.035	0.013	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.990	0.829	0.549	0.167	0.080	0.035	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	1.000	0.964	0.816	0.398	0.236	0.127	0.027	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	1.000	0.995	0.944	0.648	0.461	0.297	0.091	0.018	0.002	0.000	0.000	0.000	0.000	0.000	0.000
4	1.000	0.999	0.987	0.836	0.686	0.515	0.217	0.059	0.009	0.001	0.000	0.000	0.000	0.000	0.000
5	1.000	1.000	0.998	0.939	0.852	0.722	0.403	0.151	0.034	0.004	0.001	0.000	0.000	0.000	0.000
6	1.000	1.000	1.000	0.982	0.943	0.869	0.610	0.304	0.095	0.015	0.004	0.001	0.000	0.000	0.000
7	1.000	1.000	1.000	0.996	0.983	0.950	0.787	0.500	0.213	0.050	0.017	0.004	0.000	0.000	0.000
8	1.000	1.000	1.000	0.999	0.996	0.985	0.905	0.696	0.390	0.131	0.057	0.018	0.000	0.000	0.000
9	1.000	1.000	1.000	1.000	0.999	0.996	0.966	0.849	0.597	0.278	0.148	0.061	0.002	0.000	0.000
10	1.000	1.000	1.000	1.000	1.000	0.999	0.991	0.941	0.783	0.485	0.314	0.164	0.013	0.001	0.000
11	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.982	0.909	0.703	0.539	0.352	0.056	0.005	0.000
12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.973	0.873	0.764	0.602	0.184	0.036	0.000
13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.965	0.920	0.833	0.451	0.171	0.010
14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.987	0.965	0.794	0.537	0.140

$n = 20$

k	p														
	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.818	0.358	0.122	0.012	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.983	0.736	0.392	0.069	0.024	0.008	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.999	0.925	0.677	0.206	0.091	0.035	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	1.000	0.984	0.867	0.411	0.225	0.107	0.016	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	1.000	0.997	0.957	0.630	0.415	0.238	0.051	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	1.000	1.000	0.989	0.804	0.617	0.416	0.126	0.021	0.002	0.000	0.000	0.000	0.000	0.000	0.000
6	1.000	1.000	0.998	0.913	0.786	0.608	0.250	0.058	0.006	0.000	0.000	0.000	0.000	0.000	0.000
7	1.000	1.000	1.000	0.968	0.898	0.772	0.416	0.132	0.021	0.001	0.000	0.000	0.000	0.000	0.000
8	1.000	1.000	1.000	0.990	0.959	0.887	0.596	0.252	0.057	0.005	0.001	0.000	0.000	0.000	0.000
9	1.000	1.000	1.000	0.997	0.986	0.952	0.755	0.412	0.128	0.017	0.004	0.001	0.000	0.000	0.000
10	1.000	1.000	1.000	0.999	0.996	0.983	0.872	0.588	0.245	0.048	0.014	0.003	0.000	0.000	0.000
11	1.000	1.000	1.000	1.000	0.999	0.995	0.943	0.748	0.404	0.113	0.041	0.010	0.000	0.000	0.000
12	1.000	1.000	1.000	1.000	1.000	0.999	0.979	0.868	0.584	0.228	0.102	0.032	0.000	0.000	0.000
13	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.942	0.750	0.392	0.214	0.087	0.002	0.000	0.000
14	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.979	0.874	0.584	0.383	0.196	0.011	0.000	0.000
15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.949	0.762	0.585	0.370	0.043	0.003	0.000
16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.984	0.893	0.775	0.589	0.133	0.016	0.000
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.965	0.909	0.794	0.323	0.075	0.001
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.992	0.976	0.931	0.908	0.264	0.017
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.988	0.878	0.642	0.182

$n = 25$

k	p 0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.778	0.277	0.072	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.974	0.642	0.271	0.027	0.007	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.998	0.873	0.537	0.098	0.032	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	1.000	0.966	0.764	0.234	0.096	0.033	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	1.000	0.993	0.902	0.421	0.214	0.090	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	1.000	0.999	0.967	0.617	0.378	0.193	0.029	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	1.000	1.000	0.991	0.780	0.561	0.341	0.074	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	1.000	1.000	0.998	0.891	0.727	0.512	0.154	0.022	0.001	0.000	0.000	0.000	0.000	0.000	0.000
8	1.000	1.000	1.000	0.953	0.851	0.677	0.274	0.054	0.004	0.000	0.000	0.000	0.000	0.000	0.000
9	1.000	1.000	1.000	0.983	0.929	0.811	0.425	0.115	0.013	0.000	0.000	0.000	0.000	0.000	0.000
10	1.000	1.000	1.000	0.994	0.970	0.902	0.586	0.212	0.034	0.002	0.000	0.000	0.000	0.000	0.000
11	1.000	1.000	1.000	0.998	0.989	0.956	0.732	0.345	0.078	0.006	0.001	0.000	0.000	0.000	0.000
12	1.000	1.000	1.000	1.000	0.997	0.983	0.846	0.500	0.154	0.017	0.003	0.000	0.000	0.000	0.000
13	1.000	1.000	1.000	1.000	0.999	0.994	0.922	0.655	0.268	0.044	0.011	0.002	0.000	0.000	0.000
14	1.000	1.000	1.000	1.000	1.000	0.998	0.966	0.788	0.414	0.098	0.030	0.006	0.000	0.000	0.000
15	1.000	1.000	1.000	1.000	1.000	1.000	0.987	0.885	0.575	0.189	0.071	0.017	0.000	0.000	0.000
16	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.946	0.726	0.323	0.149	0.047	0.000	0.000	0.000
17	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.978	0.846	0.488	0.273	0.109	0.002	0.000	0.000
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.993	0.926	0.659	0.439	0.220	0.009	0.000	0.000
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.971	0.807	0.622	0.383	0.033	0.001	0.000
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.991	0.910	0.786	0.579	0.098	0.007	0.000
21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.967	0.904	0.766	0.236	0.034	0.000
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.991	0.968	0.902	0.463	0.127	0.002
23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.993	0.973	0.729	0.358	0.026
24	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.928	0.723	0.222

附录 3　标准正态分布累积概率：$P(Z \leqslant Z_{\alpha}) = \int_{-\infty}^{Z_{\alpha}} f(z)\mathrm{d}z = 1 - \alpha$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8314	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

α	0.90	0.95	0.975	0.99	0.995	0.999	0.9995	0.99995	0.999995
z_{α}	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.891	4.417

附录4 t 分布的临界值：$t_\alpha(v)P(T \geqslant t_\alpha(v)) = \int_{t_\alpha(v)}^{\infty} f(x;v)\mathrm{d}x = \alpha$

$d.f.$	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.656
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
∞	1.282	1.645	1.960	2.326	2.576

附录 5　卡方分布的临界值：$\chi_{\alpha}^{2}(v)\ P(\chi^{2} \geq \chi_{\alpha}^{2}(v)) = \int_{\chi_{\alpha}^{2}(v)}^{\infty} f(x;v)\,dx = \alpha$

d.f.	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.975}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$
1	0.0000393	0.0001571	0.0009821	0.0039322	0.0157907
2	0.0100247	0.0201004	0.0506357	0.1025862	0.2107208
3	0.0717235	0.1148316	0.2157949	0.3518460	0.5843755
4	0.206984	0.297107	0.484419	0.710724	1.063624
5	0.411751	0.554297	0.831209	1.145477	1.610309
6	0.675733	0.872083	1.237342	1.635380	2.204130
7	0.989251	1.239032	1.689864	2.167349	2.833105
8	1.344403	1.646506	2.179725	2.732633	3.489537
9	1.734911	2.087889	2.700389	3.325115	4.168156
10	2.155845	2.558199	3.246963	3.940295	4.865178
11	2.603202	3.053496	3.815742	4.574809	5.577788
12	3.073785	3.570551	4.403778	5.226028	6.303796
13	3.565042	4.106900	5.008738	5.891861	7.041500
14	4.074659	4.660415	5.628724	6.570632	7.789538
15	4.600874	5.229356	6.262123	7.260935	8.546753
16	5.142164	5.812197	6.907664	7.961639	9.312235
17	5.697274	6.407742	7.564179	8.671754	10.0852
18	6.264766	7.014903	8.230737	9.390448	10.8649
19	6.843923	7.632698	8.906514	10.1170	11.6509
20	7.433811	8.260368	9.590772	10.8508	12.4426
21	8.033602	8.897172	10.2829	11.5913	13.2396
22	8.642681	9.542494	10.9823	12.3380	14.0415
23	9.260383	10.1957	11.6885	13.0905	14.8480
24	9.886199	10.8563	12.4011	13.8484	15.6587
25	10.5196	11.5240	13.1197	14.6114	16.4734
26	11.1602	12.1982	13.8439	15.3792	17.2919
27	11.8077	12.8785	14.5734	16.1514	18.1139
28	12.4613	13.5647	15.3079	16.9279	18.9392
29	13.1211	14.2564	16.0471	17.7084	19.7677
30	13.7867	14.9535	16.7908	18.4927	20.5992
40	20.7066	22.1642	24.4331	26.5093	29.0505
50	27.9908	29.7067	32.3574	34.7642	37.6886
60	35.5344	37.4848	40.4817	43.1880	46.4589
80	51.1719	53.5400	57.1532	60.3915	64.2778
100	67.3275	70.0650	74.2219	77.9294	82.3581

续 表

d.f.	$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.010}$	$\chi^2_{0.005}$
1	2.705541	3.841455	5.023903	6.634891	7.879400
2	4.605176	5.991476	7.377779	9.210351	10.5965
3	6.251394	7.814725	9.348404	11.3449	12.8381
4	7.779434	9.487728	11.1433	13.2767	14.8602
5	9.236349	11.0705	12.8325	15.0863	16.7496
6	10.6446	12.5916	14.4494	16.8119	18.5475
7	12.0170	14.0671	16.0128	18.4753	20.2777
8	13.3616	15.5073	17.5345	20.0902	21.9549
9	14.6837	16.9190	19.0228	21.6660	23.5893
10	15.9872	18.3070	20.4832	23.2093	25.1881
11	17.2750	19.6752	21.9200	24.7250	26.7569
12	18.5493	21.0261	23.3367	26.2170	28.2997
13	19.8119	22.3620	24.7356	27.6882	29.8193
14	21.0641	23.6848	26.1189	29.1412	31.3194
15	22.3071	24.9958	27.4884	30.5780	32.8015
16	23.5418	26.2962	28.8453	31.9999	34.2671
17	24.7690	27.5871	30.1910	33.4087	35.7184
18	25.9894	28.8693	31.5264	34.8052	37.1564
19	27.2036	30.1435	32.8523	36.1908	38.5821
20	28.4120	31.4104	34.1696	37.5663	39.9969
21	29.6151	32.6706	35.4789	38.9322	41.4009
22	30.8133	33.9245	36.7807	40.2894	42.7957
23	32.0069	35.1725	38.0756	41.6383	44.1814
24	33.1962	36.4150	39.3641	42.9798	45.5584
25	34.3816	37.6525	40.6465	44.3140	46.9280
26	35.5632	38.8851	41.9231	45.6416	48.2898
27	36.7412	40.1133	43.1945	46.9628	49.6450
28	37.9159	41.3372	44.4608	48.2782	50.9936
29	39.0875	42.5569	45.7223	49.5878	52.3355
30	40.2560	43.7730	46.9792	50.8922	53.6719
40	51.8050	55.7585	59.3417	63.6908	66.7660
50	63.1671	67.5048	71.4202	76.1538	79.4898
60	74.3970	79.0820	83.2977	88.3794	91.9518
80	96.5782	101.879	106.629	112.329	116.321
100	118.498	124.342	129.561	135.807	140.170

附录 6　F 分布的临界值：$F_\alpha(v_1,v_2)P(F \geqslant F_\alpha(v_1,v_2)) = \int_{F_\alpha(v_1,v_2)}^{\infty} f(x;v_1,v_2)\mathrm{d}x = \alpha$

$v_2(d.f.)$	$v_1(d.f.)$　$\alpha = 0.10$								
	1	2	3	4	5	6	7	8	9
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

续 表

$\nu_2(d.f.)$	$\nu_1(d.f.)$ $\alpha = 0.10$									
	10	12	15	20	24	30	40	60	120	∞
1	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.11
6	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

续　表

$\nu_2(d.f.)$	$\nu_1(d.f.)$　$\alpha=0.05$								
	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

续 表

$v_2(d.f.)$	$v_1(d.f.)$ $\alpha=0.05$									
	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.90	245.95	248.02	249.05	250.10	251.14	252.20	253.25	254.32
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

续　表

$v_2(d.f.)$	$v_1(d.f.)$ $\alpha=0.025$								
	1	2	3	4	5	6	7	8	9
1	647.79	799.48	864.15	899.60	921.83	937.11	948.20	956.64	963.28
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11

续 表

$v_2(d.f.)$	$v_1(d.f.)$ $\alpha = 0.025$									
	10	12	15	20	24	30	40	60	120	∞
1	968.63	976.72	984.87	993.08	997.27	1001.40	1005.60	1009.79	1014.04	1018.26
2	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	4.76	4.67	4.57	4.47	4.41	4.36	4.31	4.25	4.20	4.14
8	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

附录 7　威尔克森符号秩检验概率

ONE-TAILED	TWO-TAILED	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	$n=11$	$n=12$
$\alpha=0.05$	$\alpha=0.10$	1	2	4	6	8	11	14	17
$\alpha=0.025$	$\alpha=0.05$		1	2	4	6	8	11	14
$\alpha=0.01$	$\alpha=0.02$			0	2	3	5	7	10
$\alpha=0.005$	$\alpha=0.01$				0	2	3	5	7

ONE-TAILED	TWO-TAILED	$n=13$	$n=14$	$n=15$	$n=16$	$n=17$	$n=18$	$n=19$	$n=20$
$\alpha=0.05$	$\alpha=0.10$	21	26	30	36	41	47	54	60
$\alpha=0.025$	$\alpha=0.05$	17	21	25	30	35	40	46	52
$\alpha=0.01$	$\alpha=0.02$	13	16	20	24	28	33	38	43
$\alpha=0.005$	$\alpha=0.01$	10	13	16	19	23	28	32	37

ONE-TAILED	TWO-TAILED	$n=21$	$n=22$	$n=23$	$n=24$	$n=25$	$n=26$	$n=27$	$n=28$
$\alpha=0.05$	$\alpha=0.10$	68	75	83	92	101	110	120	130
$\alpha=0.025$	$\alpha=0.05$	59	66	73	81	90	98	107	117
$\alpha=0.01$	$\alpha=0.02$	49	56	62	69	77	85	93	102
$\alpha=0.005$	$\alpha=0.01$	43	49	55	61	68	76	84	92

ONE-TAILED	TWO-TAILED	$n=29$	$n=30$	$n=31$	$n=32$	$n=33$	$n=34$	$n=35$	$n=36$
$\alpha=0.05$	$\alpha=0.10$	141	152	163	175	188	201	214	228
$\alpha=0.025$	$\alpha=0.05$	127	137	148	159	171	183	195	208
$\alpha=0.01$	$\alpha=0.02$	111	120	130	141	151	162	174	186
$\alpha=0.005$	$\alpha=0.01$	100	109	118	128	138	149	160	171

ONE-TAILED	TWO-TAILED	$n=37$	$n=38$	$n=39$	$n=40$	$n=41$	$n=42$	$n=43$	$n=44$
$\alpha=0.05$	$\alpha=0.10$	242	256	271	287	303	319	336	353
$\alpha=0.025$	$\alpha=0.05$	222	235	250	264	279	295	311	327
$\alpha=0.01$	$\alpha=0.02$	198	211	224	238	252	267	281	297
$\alpha=0.005$	$\alpha=0.01$	183	195	208	221	234	248	262	277

ONE-TAILED	TWO-TAILED	$n=45$	$n=46$	$n=47$	$n=48$	$n=49$	$n=50$
$\alpha=0.05$	$\alpha=0.10$	371	389	408	427	446	466
$\alpha=0.025$	$\alpha=0.05$	344	361	379	397	415	434
$\alpha=0.01$	$\alpha=0.02$	313	329	345	362	380	398
$\alpha=0.005$	$\alpha=0.01$	292	307	323	339	356	373

附录 8　威尔克森秩和检验概率

A. $\alpha = 0.025$ one-tailed; $\alpha = 0.05$ two-tailed

$n_2 \setminus n_1$	3		4		5		6		7		8		9		10	
	W_L	W_U	W_L	W_U	W_L	W_U	W_L	W_U	W_L	W_U	W_L	W_U	W_L	W_U	W_L	W_U
3	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33
4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44
5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56
6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70
7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83
8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

B. $\alpha = 0.05$ one-tailed; $\alpha = 0.10$ two-tailed

$n_2 \setminus n_1$	3		4		5		6		7		8		9		10	
	W_L	W_U	W_L	W_U	W_L	W_U	W_L	W_U	W_L	W_U	W_L	W_U	W_L	W_U	W_L	W_U
3	6	15	7	17	7	20	8	22	9	24	9	27	10	29	11	31
4	7	17	12	24	13	27	14	30	15	33	16	36	17	39	18	42
5	7	20	13	27	19	36	20	40	22	43	24	46	25	50	26	54
6	8	22	14	30	20	40	28	50	30	54	32	58	33	63	35	67
7	9	24	15	33	22	43	30	54	39	66	41	71	43	76	46	80
8	9	27	16	36	24	46	32	48	41	71	52	84	54	90	57	95
9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	69	111
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

n	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
5	0.900	—	—	—
6	0.829	0.886	0.943	—
7	0.714	0.786	0.893	—
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.523	0.623	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.564	0.625
19	0.388	0.462	0.549	0.608
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562
23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.537
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.377	0.448	0.496
29	0.311	0.370	0.440	0.487
30	0.305	0.364	0.432	0.478

参 考 文 献

中文部分

[1] 吴柏林.模糊统计分析:问卷调查研究的新方向[J].中国台湾政治大学研究通讯,1995,2:65-80.

[2] 吴柏林.社会科学研究中的模糊逻辑与模糊统计分析[J].中国台湾政治大学研究通讯,1997,7:17-38.

[3] 吴柏林,许毓云.模糊统计分析在台湾地区失业率应用[J].中国统计学报,1999,37(1):37-52.

[4] 吴柏林,曾能芳.模糊回归参数估计及在景气对策信号之分析应用[J].中国统计学报,1998,36(4):399-420.

[5] 吴柏林,张钿富,廖敏治.模糊时间数列与台湾地区中学教师需求人数之预测[J].中国台湾政治大学学报,1996,73:287-312.

[6] 李允中,王小璠,苏木春.模糊理论及其应用[M].台北:全华科技图书股份有限公司,2003.

[7] 林原宏,郑舜仁,吴柏林.模糊众数及其在教育与心理评量分析之应用[J].中国统计学报,2003,41(1):39-66.

[8] 胡悦伦,吴柏林.模糊统计在分析乐观量表之应用[J].教育与心理研究,2002,25:457-484.

[9] 张钿富,吴柏林.台湾地区中小学在学人数结构改变之探讨[J].中国台湾编译馆馆刊,1996,25(2):247-262.

[10] 陈国任,林雅惠,吴柏林,等.模糊统计分析及在茶叶质量评定的应用[J].台湾茶叶研究汇报,1998,17:19-37.

[11] 叶秋呈,施耀振,吴柏林.应用模糊众数与模糊期望值于大学生多元学习生活调查分析[J].智慧科技与应用统计学报,2004,2(1):109-136.

[12] 陈宇煌.系统模糊决策理论与应用[M].大连:大连理工大学出版社,1994.

[13] 邓成梁.运筹学的原理和方法[M].武汉:华中理工大学出版社,1996.

[14] 冯德益,楼世博.模糊数学:方法与应用[M].台北:科技图书股份有限公司,1991.

[15] 郭桂蓉.模糊模式识别[M].长沙:国防科技大学出版社,1993.

[16] 胡玉立,等.市场预测与管理决策[M].北京:中国人民大学出版社,1996.

[17] 李安宁,吴达.模糊数学基础及其应用[M].乌鲁木齐:新疆人民出版社,1986.
[18] 李怀祖.决策理论导引[M].北京:机械工业出版社,1992.
[19] 汤兵勇,王文杰,郑飞.现代模糊管理数学方法[M].上海:中国纺织大学出版社,1999.
[20] 吴秉坚.模糊数学及其经济分析[M].北京:中国标准出版社,1994.
[21] 吴达,吴柏林.模糊回归参数估计方法与应用[J].系统工程理论实践,2001,11:61-67.
[22] 吴柏林,林玉钧.模糊时间数列分析与预测:以台湾地区加权股价指数为例[J].应用数学学报,2002,25(1):67-76,
[23] 孙山泽.非参数统计学讲义[M].北京:北京大学出版社,2000.
[24] 张跃,王光远.模糊随机动力系统理论[M].北京:科学出版社,1993.
[25] 李鸿吉.模糊数学基础及实用算法[M].北京:科学出版社,2005.
[26] 韩立岩,汪培庄.应用模糊数学[M].修订版.北京:首都经济贸易大学科学出版社,1998.
[27] 陈水利,李敬功,王向公.模糊集理论及其应用[M].北京:科学出版社,2005.
[28] 曹炳元.应用模糊数学与系统[M].北京:科学出版社,2005.
[29] 杨志民,刘广利.不确定性支持向量机原理及应用不可失[M].北京:科学出版社,2007.
[30] 邓乃杨,田英杰.数据挖掘中的新方法——支持向量机[M].北京:科学出版社,2004.

英文部分

[1] ABBOUD N J, SAKAWA M, INUIGUCHI M. A Fuzzy Programming Approach to Multiobjective Multidimensional 0-1 Knapsack Problems[J]. Fuzzy Sets and Systems, 1997, 86:1-14.
[2] ALBERTI N, PERRONE G. Multipass Machining Optimization by Using Fuzzy Possibilistic Programming and Genetic Algorithms[J]. Proceedings of the Institution of Mechanical Engineers, 1999, 213(3):273.
[3] ALUJA J. Towards a New Paradigm of Investment Selection in Uncertainty[J]. Fuzzy Sets and Systems, 1996, 84:187-197.
[4] AMO A, MONTERO J, BIGING G, et al. Fuzzy Classification Systems[J]. European Journal of Operational Research, 2004, 156(2):495.
[5] BAGNOLI C, SMITH H C. The Theory of Fuzzy Logic and Its Application to Real

Estate Valuation[J]. Journal of Real Estate Research, 1998, 16(2):169-199.

[6] BALDWIN J, LAWRY J, MARTIN T. The Application of Generalized Fuzzy Rules to Machine Learning and Automated Knowledge Discovery[J]. Inter. J. Uncertainty, Fuzziness and Knowledge-Based Systems, 1998, 6(5):459-487.

[7] BANDERNER H, NATHER W. Fuzzy Data Analysis[M]. Dordrecht: Kluwer Academic, 1992.

[8] BARALDI A, BLONDA P. A Survey of Fuzzy Clustering Algorithms for Pattern Recognition. Part I and Part II[J]. IEEE Trans. Syst., Man, Cybern. Part B: CYBERN, 1999, 29(6):778-801.

[9] BOREUX J J, HENRY C G, GUIOT J, et al. Radial Tree-growth Modeling with Fuzzy Regression[J]. Canadian Journal of Forest Research, 1998, 28(8):1249-1260.

[10] BUCKLEY J J. Fuzzy Probabilities: New Approach and Applications[M]. Berlin: Physics-Verlag, 2003.

[11] BUCKLEY J J. Fuzzy Statistics[M]. Heidelberg: Springer-Verlag, 2004.

[12] BUFARDI A. Fuzzy Sets in Decision Analysis, Operations Research and Statistics [J]. Journal of Multicriteria Decision Analysis, 1998, 8(2):112.

[13] CARRANO A L, TAYLOR J B, YOUNG R E, et al. Fuzzy Knowledge-based Modeling and Statistical Regression in Abrasive Wood Machining[M]. Forest Products Journal, 2004, 54(5):66-72.

[14] CASALINO G, MEMOLA F, MINUTOLO C. A Model for Evaluation of Laser welding Efficiency and Quality Using an Artificial Neural Network and Fuzzy Logic [J]. Proceedings of the Institution of Mechanical Engineers, 2004, 218(6):641.

[15] CASALS M R, GIL P. Bayesian Sequential Test for Fuzzy Parametric Hypotheses from Fuzzy Information[J]. Information Sciences, 1994, 80:283-298.

[16] CASALS M R, GIL M A, GIL P. The Fuzzy Decision Problem: An Approach to the Problem of Testing Statistical Hypotheses with Fuzzy Information[J]. European Journal of Operational Research, 1986, 27:371-382.

[17] CHANG P T. Fuzzy Stage Characteristic-preserving Product Life Cycle Modeling [J]. Fuzzy Sets and Systems, 2002, 126 (1):33-47.

[18] CHANG Y O, AYYUB B M. Fuzzy Regression Methods—A Comparative Assessment[J]. Fuzzy Sets and Systems, 2001, 119:187-203.

[19] CHEN S R, WU B. On Optimal Forecasting with Soft Computation for Nonlinear Time Series[J]. Fuzzy Optimization and Decision Making, 2003, 2(3):215.

[20] CHEN C, CHERN T, WU B. On the Selection of Subset Bilinear Time Series Models: A Genetic Algorithm Approach[J]. Computational Statistics, 2001, 16: 505-517.

[21] CHEN L H, KAO C, KUO S, et al. Productivity Diagnosis via Fuzzy Clustering and Classification: An Application to Machinery Industry[J]. Omega, Int. J. Mgmt Sci., 1996, 24(3):309-319.

[22] CHEN L S, CHENG C H. Selecting IS Personnel Use Fuzzy GDSS Based on Metric Distance Method[J]. European Journal of Operational Research, 2005, 160: 803-820.

[23] CHEN S, WU B. On Optimal Forecasting with Soft Computation for Nonlinear Time Series[J]. Fuzzy Optimization and Decision Making, 2003, 3:215-228.

[24] CHEN S M. Forecasting Enrollments Based on Fuzzy Time Series[J]. Fuzzy Sets and Systems, 1996, 81:311-319.

[25] CHENG C H. A New Approach for Ranking Fuzzy Numbers by Distance Method [J]. Fuzzy Sets and Systems, 1998, 95:307-317.

[26] CHENG T W, GOLDGOF D B, HALL L O. Fast Fuzzy Clustering[J]. Fuzzy Sets and Systems, 1998, 93:49-56.

[27] CHENG T, MOLENAAR M, LIN H. Formalizing Fuzzy Objects from Uncertain Classification Results[J]. International Journal of Geographical Information Science, 2001, 15(1):27-42.

[28] CHIANG D A, CHOW L R, WANG Y F. Mining Time Series Data by a Fuzzy Linguistic Summary System[J]. Fuzzy Sets and Systems, 2000, 112:419-432.

[29] CHIANG W Y, LIANG G S, YAHALOM S Z. The Fuzzy Clustering Method: Applications in the Air Transport Market in Taiwan[J]. Journal of Database Marketing & Customer Strategy Management, 2003, 11(2):149.

[30] CLYMER J, COREY P, GARDNER J. Discrete Event Fuzzy Airport Control[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1992, 22(2):343-351.

[31] COPPI R. A Theoretical Framework for Data Mining: The "Information Paradigm" [J]. Computational Statistics & Data Analysis, 2002, 38:501-515.

[32] CUSTEM B V, GATH I. Detection of Outliers and Robust Estimation Using Fuzzy Clustering[J]. Computational Statistics and Data Analysis, 1993, 15:47-61.

[33] DIAMOND P, KLODEN P. Metric Spaces of Fuzzy Sets [M]. Singapore: World Scientific, 1994.

[34] DIAMOND P. Fuzzy Least Squares[J]. Information Science, 1988, 46:141-157.

[35] DONG H, PARK D. An Efficient Algorithm for Fuzzy Weighted Average[J]. Fuzzy Sets And Systems, 1997, 87:39-45.

[36] DORSEY D W, COOVERT M D. Mathematical Modeling of Decision Making: A Soft and Fuzzy Approach to Capturing Hard Decisions[J]. Human Factors, 2003, 45(1):117.

[37] DRIANKOV D, HELLENDOORN H, REINFRANK M. An Introduction to Fuzzy Control[M]. New York: Springer-Verlag, 1996.

[38] DUBOIS D, PRADE H. Fuzzy Sets in Approximate Reasoning, Part 1: Inference with Possibility Distributions[J]. Fuzzy Sets and Systems, 1991, 40:143-202.

[39] ESOGBUE A O, SONG Q. On Optimal Defuzzification and Learning Algorithms: Theory and Applications[J]. Fuzzy Optimization and Decision Making, 2003, 2(4):283.

[40] GALVO T, MESIAR R. Generalized Medians[J]. Fuzzy Sets and Systems, 2001, 124:59-64.

[41] GEHRKE M, WALKER C, WALKER E. A Mathematical Setting for Fuzzy Logic [J]. Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems, 1997, 5 (3):223-228.

[42] GERTNER G Z, ZHU H. Bayesian Estimation in Forest Surveys When Samples or Prior Information are Fuzzy[J]. Fuzzy Sets and Systems, 1996, 77:277-290.

[43] GIL M A, CORRAL N, GIL P. The Minimum Inaccuracy Estimates in χ^2 Tests for Goodness of Fit with Fuzzy Observations[J]. Journal of Statistical Planning and Inference, 1998, 19:95-115.

[44] GORSEVSKI P V, GESSLER P E, JANKOWSK P. Integrating a Fuzzy k-means Classification and a Bayesian Approach for Spatial Prediction of Landslide Hazard [J]. Journal of Geographical Systems, 2003, 5(3): 223.

[45] GUARISO G, RIZZOLI A, WERTHNER H. Identification of Model Structure via Qualitive Simulation[J]. IEEE Trans. on Systems, Man, and Cybernetics, 1992, 22(5):1075-1086.

[46] GVISHIANI A D, AGAYAN S M, BOGOUTDINOV S R, et al. Mathematical Methods of Geoinformatics. II. Fuzzy-Logic Algorithms in the Problems of Abnormality Separation in Time Series[J]. Cybernetics and Systems Analysis, 2003, 39(4):555.

[47] HAGGSTROM O. Positive Correlations in the Fuzzy Potts Model[J]. Annals of Applied Probability, 1999, 9(4):1149-1159.

[48] HAN J S, BANG W C, BIEN Z Z. Feature Set Extraction Algorithm Based on Soft Computing Techniques and Its Application to EMG Pattern Classification[J]. Fuzzy Optimization and Decision Making, 2002, 1(3):269.

[49] HARRIS T, STODDARD S, BEZDEK J. Application of Fuzzy-set Clustering for Regional Typologies[J]. Growth and Change, 1993, 24:155-165.

[50] HATHAWAY R J, BEZDEK J C. Switching Regression Models and Fuzzy Clustering[J]. IEEE Transactions of Fuzzy Systems, 1993, 1:195-204.

[51] HOHLE U, KLEMENT P. Non-Classical Logics and Their Applications to Fuzzy Subsets: A Handbook of the Mathematical Foundations Of Fuzzy Set Theory[M]. Boston: Kluwer Academic, 1995.

[52] HONG T, LEE C. Induction of Fuzzy Rules and Membership Functions from Training Examples[J]. Fuzzy Sets and Systems, 1996, 84:33-47.

[53] HöPPNER F, KLAWONN F, KRUSE R, et al. Fuzzy Cluster Analysis[M]. New York: John Wiley &Son, LTD., 1999.

[54] HORIA F, SARBU C. A New Fuzzy Regression Algorithm[J]. Analytical Chemistry, 1996, 68(5):771.

[55] HU Y C, CHEN R S, TZENG G H. Mining Fuzzy Association Rules for Classification Problems[J]. Computers & Industrial Engineering, 2002, 43(4):735.

[56] HUARNG K. Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series[J]. Fuzzy Sets and Systems, 2001, 123:387-394.

[57] HUARNG K. Heuristic Models of Fuzzy Time Series for Forecasting[J]. Fuzzy Sets and Systems, 2001, 123: 369-386.

[58] HUNG W L, WU J W. A Note on the Correlation of Fuzzy Numbers by Expected Interval[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2001, 9:517-523.

[59] HUNG W L, WU J W. Correlation of Fuzzy Numbers by α-cut Method[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2000, 10:725-735.

[60] HWANG C M, YAO J S. Independent Fuzzy Random Variables and Their Application[J]. Fuzzy Sets and Systems, 1996, 82:335-350.

[61] HWANG J R, CHEN S M, LEE C H. Handling Forecasting Problems Using Fuzzy Time Series[J]. Fuzzy Sets and Systems, 1998, 100:217-228.

[62] IANCU I. Reasoning System with Fuzzy Uncertainty[J]. Fuzzy Sets and Systems, 1997, 9:251-259.

[63] IPPOLITO L, LOIA V, SIANO P. Extended Fuzzy C-Means and Genetic Algorithms to Optimize Power Flow Management in Hybrid Electric Vehicles[J]. Fuzzy Optimization and Decision Making, 2003, 2(4):359.

[64] ISHIBUCHI H, YAMAMOTO T. Comparison of Heuristic Criteria for Fuzzy Rule Selection in Classification Problems[J]. Fuzzy Optimization and Decision Making, 2004, 3(2):119.

[65] JANG J-S, SUN C T, MIZUTANI E. Neuro-fuzzy and Soft-computing[M]. New York:Pretice-Hall International, Inc, 1997.

[66] JAULENT M C, JOYAUX C, COLOMBET I, et al. Modeling Uncertainty in Computerized Guidelines Using Fuzzy Logic[J]. Journal of the American Medical Informatics Association, 2001, 8(4):284-288.

[67] JOO Y, HWANG H, KIM K, et al. Fuzzy System Modeling by Fuzzy Partition and GA Hybrid Schemes[J]. Fuzzy Sets and Syst, 1997, 86:279-288.

[68] JORQUERA H, PALMA W, TAPIA J. A Ground-level Ozone Forecasting Model for Santiago, Chile[J]. Journal of Forecasting, 2002, 21(6):451.

[69] KAO C, CHYU C L. Least-squares Estimates in Fuzzy Regression Analysis[J]. European Journal of Operational Research, 2003, 148(2):426.

[70] KAO C, CHYU C L. A Fuzzy Linear Regression Model with Better Explanatory Power[J]. Fuzzy Sets and Systems, 2002, 126: 401-409.

[71] KIM K J, MOSKOWITZ H, DHINGRA A, et al. Fuzzy Multicriteria Models for Quality Function Deployment [J]. European Journal of Operational Research, 2000, 121 (3):504-518.

[72] KLEYLE R, KORVIN A, MCLAUGHLIN T. Decision Making on the Basis of Expected Cost Variance: A Fuzzy Set Approach [J]. Managerial Finance, 1996, 22(11):18-29.

[73] KLIR G, YUAN. Fuzzy Sets and Fuzzy Logic-Theory and Applications[M]. Upper Saddle River: Prentice-Hall, 1995.

[74] KORNER R. On the Variance of Fuzzy Random Variables[J]. Fuzzy Sets and Systems, 1997, 92:83-93.

[75] KOSKO B. Fuzzy Thinking:The New Science of Fuzzy Logic[M]. New York:Hyperion, 1994.

[76] KREINOVICH V, NGUYEN H, WU B. Justification of Heuristic Methods in Data Processing Using Fuzzy Theory, with Applications to Detection of Business Cycles from Fuzzy Data[M]. East-West Journal of Mathematics, 1999, 1(2):147-157.

[77] KUMAR K, WU B. Detection of Change Points in Time Series Analysis with Fuzzy Statistics[J]. International Journal of Systems Science, 2001, 32(9): 1185- 1192.

[78] KUO C F, SHIH C Y, LEE J Y. Automatic Recognition of Fabric Weave Patterns by a Fuzzy C-Means Clustering Method[J]. Textile Research Journal, 2004, 74 (2):107.

[79] KUO R J. A Sales Forecasting System Based on Fuzzy Neural Network with Initial Weights Generated by Genetic Algorithm[J]. European Journal of Operational Research, 2001, 129(3):496.

[80] KUO R J, XUE K C. A Decision Support System for System Sales Forecasint through Fuzzy Neural Networks with Asymmetric Fuzzy Weights[J]. Decision Support Systems, 1998, 24(2):105-126.

[81] PARK K S, KIM S H. A Note on the Fuzzy Weighted Additive Rule[J]. Fuzzy Sets and Systems, 1996, 77:315-320.

[82] LEE T S, CHIU C C, LIN F C. Prediction of the Unemployment Rate Using Fuzzy Time Series with Box-Jenkins Methodology [J]. International Journal of Fuzzy Systems, 2001, 3(4):577-585.

[83] LEE H S, CHOU M T. Fuzzy Forecasting Based on Fuzzy Time Series[J]. International Journal of Computer Mathematics, 2004, 81(7):781-789.

[84] LEE Y C, HWANG C, SHIH Y P. A Combined Approach to Fuzzy Model Identification[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1994, 24 (5):736-743.

[85] LIN C C, CHEN A P. Fuzzy Discriminant Analysis with Outlier Detection by Genetic Algorithm[J]. Computers & Operations Research, 2004, 31(6): 877.

[86] LINGRAS P, WEST C. Interval Set Clustering of Web Users with Rough K-Means [J]. Journal of Intelligent Information Systems, 2004, 23(1):5.

[87] LIU S T, KAO C. Fuzzy Measures for Correlation Coefficient of Fuzzy Numbers [J]. Fuzzy Sets and Systems, 2002, 128:267-275.

[88] LIU W Y, SONG N. The Fuzzy Association Degree in Semantic Data Models[J]. Fuzzy Sets and Systems, 2001, 117:203-208.

[89] LIU Y K, LIU B. Fuzzy Random Variables: A Scalar Expected Value Operator [J]. Fuzzy Optimization and Decision Making, 2003, 2(2):143.

[90] LOWEN R. A Fuzzy Language Interpolation Theorem[J]. Fuzzy Sets and Systems, 1990, 34:33-38.

[91] LUCIEER A, KRAAK M J. Interactive and Visual Fuzzy Classification of Remotely Sensed Imagery for Exploration of Uncertainty[J]. International Journal of Geographical Information Science, 2004, 18 (5):491-512.

[92] MANTON K. WOODBURY K, TOLLEY H. Statistical Applications - Using Fuzzy Set[M]. New York:John Willy & Sons, Inc, 1993.

[93] MEYEROWITZ A, RICHMAN F, WALKER E A. Calculating Maximum Entropy Densities for Belief Functions[J]. Intern. J. Uncertainty, Fuzziness and Knowledge-Based Systems, 1994, 2(4):377-389.

[94] MILAN M. Computation over Fuzzy Quantities[M]. Boca Raton: CRC Press, 1994.

[95] MMTON K G, WOODBUY M A, TOLLEY H D. Statistical Applications Using Fuzzy Sets[M]. New York: Wiley, 1994.

[96] NARAZAKI H, RALESCU A. Iterative Induction of a Category Membership Function[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1994, 2(1):91-100.

[97] NGUYEN H, WU B. Fuzzy Mathematics and Statistical Applications [M]. Taipei:Hua-Tai Book Company, 2000.

[98] NGUYEN H, WANG T, WU B. On Probabilistic Methods in Fuzzy Theory[J]. International Journal of Intelligent Systems, 2004, 19:99-109.

[99] NGUYEN H. Fuzzy Sets and Probability[J]. Fuzzy Sets and Systems, 1997, 90: 129-132.

[100] NGUYEN H, SUGENO M. Fuzzy Modeling and Control[M]. Boca Raton: CRC Press, 1998.

[101] NGUYEN H, KREINOVICH V. Applications of Continuous Mathematics to Computer Science[M]. Dorderecht: Kluwer Academic Publishers, 1997.

[102] NGUYEN H, WU B. Fuzzy/Probability ~ Fractal/Smooth [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1999, 7(4): 363-370.

[103] NGUYEN H T, PRASAD R. Fuzzy Modeling and Control-Selected Works of M. Sugeno[M]. Boca Raton: CRC Press, 1999.

[104] NGUYEN H T, WALKER E A. A First Course in Fuzzy Logic[M]. 2nd ed. Boca Raton:Chapman and Hall/CRC, 1999.

[105] NGUYEN H T, KREINOVICH V, KOSHELEVA A. Is the Success of Fuzzy Logic Really Paradoxical? [J]. Inter. J. Intelligent Systems, 1996, (5):295-

326.
[106] NGUYEN H, WU B, KREINOVICH V. On Combining Statistical and Fuzzy Techniques: Detection of Business Cycles from Uncertain Data[J]. In Trends in Information Technology, 2000, 1:69-74.
[107] NIE J. Nonlinear Time Series Forecasting: A Fuzzy-neural Approach[J]. Neurocompting, 1997, 16:63-76.
[108] ORLOWSKI S A. Calculus of Decomposable Properties, Fuzzy Sets and Decisions[M]. New York: Allerdon Press, INC, 1994.
[109] OZELKAN E C, DUCKSTEIN L. Multi-objective Fuzzy Regression: A General Framework[J]. Computers & Operations Research, 2000, 27(7, 8): 635.
[110] PAL N, BEZDEK J. On Cluster Validity for the Fuzzy c-means Model[J]. IEEE Trans. Fuzzy Systems, 1995, 3(3):370-379.
[111] PALIWAL R, GEEVARGHESE G A, BABU P R, et al. Valuation of Landmass Degradation Using Fuzzy Hedonic Method: A Case Study of National Capital Region[J]. Environmental & Resource Economics, 1999, 14 (4):519-543.
[112] PALM R, DDMIKOV D, HELLENDOORN H. Model-Based Fuzzy Control[M]. Berlin: Springer-Verlag, 1997.
[113] PARK K, KIM S. A Note on the Fuzzy Weighted Additive Rule[J]. Fuzzy Sets and Systems, 1996, 77:315-320.
[114] PARK Y M, MOON U C, LEE K Y. A Self-organizing Fuzzy Logic Controller for Dynamic Systems Using Fuzzy Auto-regressive Moving Average(FARMA) Model [J]. IEEE Transactions on Fuzzy Systems, 1995, 3(1):75-82.
[115] PETERS G. A Linear Forecasting Model and Its Application to Economic Data [J]. Journal of Forecasting, 2001, 20(5):315.
[116] RAYWARD-SMITH V J. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition[J]. The Journal of the Operational Research Society, 2000, 51(6):769.
[117] ROBERGE P. A Knowledge-based Framework for Planning Non-destructive Evaluation[J]. Expert Systems, 1995, 12(2):107-112.
[118] ROMER C, KANDEL A, BACKER E. Fuzzy Partitions of the Sample Space and Fuzzy Parameter Hypotheses[J]. IEEE Trans. Systems, Man and Cybernet. 1995, 25(9):1314-1321.
[119] RONALD R Y. Constrained OWA Aggregation[J]. Fuzzy Sets and Systems, 1996, 81:89-101.

[120] RONALD C, BLAIR J. Designing Surveys-A Guide to Decisions and Procedures [M]. Thousand Oaks: Pine Forge, 1996.

[121] ROSS T. Fuzzy Logic with Engineering Applications [M]. New York: McGraw-Hill, 1995.

[122] RUAN D. Fuzzy Set Theory and Advanced Mathematical Application [M]. Boston: Kluwer Academic, 1995.

[123] RUNKLER T A. Selection of Appropriate Defuzzification Methods Using Application Specific Properties[J]. IEEE Transactions on Fuzzy Systems, 1997, 5(1): 72-79.

[124] RUSPINI E. Approximate Reasoning: Past, Present, Future[J]. Information Sciences, 1991, 57:297-317.

[125] SANCHEZ J D, GOMEZ A T. Estimating a Fuzzy Term Structure of Interest Rates Using Fuzzy Regression Techniques[J]. European Journal of Operational Research, 2004, 154 (3):804-818.

[126] SANCHEZ J D, GOMEZ A T. Applications of Fuzzy Regression in Actuarial Analysis[J]. Journal of Risk and Insurance, 2003, 70 (4):665-699.

[127] SANCHEZ J D, GOMEZ A T. Estimating a Term structure of Interest Rates for Fuzzy Financial Pricing by Using Fuzzy Regression Methods[J]. Fuzzy Sets and Systems, 2003, 139 (2):313-331.

[128] SATO M, SATO Y. On a Multicriteria Fuzzy Clustering Method for 3-way Data [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1994, 2:127-142.

[129] SONG Q, CHISSOM B S. Forecasting Enrollments with Fuzzy Time series-Part I [J]. Fuzzy Sets and Systems, 1993, 54:1-9.

[130] SONG Q, CHISSOM B S. Fuzzy Time Series and Its Models[J]. Fuzzy Sets and Systems, 1993, 54:269-277.

[131] SONG Q, CHISSOM B S. Forecasting Enrollments with Fuzzy Time Series-Part II [J]. Fuzzy Sets and Systems, 1994, 62:1-8.

[132] SONG Q, LELAND R P. Adaptive Learning Defuzzification Techniques and Application[J]. Fuzzy Sets and Systems, 1996, 81:321-329.

[133] SONG Q, LELAND R P, CHISSOM B S. Fuzzy Stochastic Fuzzy Time Series and Its Models[J]. Fuzzy Sets and Systems, 1997, 88:333-341.

[134] SQUIRE W, TRAPP G. Using Complex Variables to Estimate Derivatives of Real Functions[J]. SIAM Reviews, 1998, 40(1):110-112.

[135] STOJAKOVIC M. Fuzzy Conditional Expectation[J]. Fuzzy Sets and Systems, 1992, 52:53-60.

[136] STOJAKOVIC M. Fuzzy Random Variables, Expectation, and Martingales[J]. Journal of Mathematical Analysis and Applications, 1994, 184:594-606.

[137] SULLIVAN J, WOODALL W H. Comparison of Fuzzy Forecasting and Markov Modeling[J]. Fuzzy Sets and Systems, 1994, 64:279-293.

[138] SWAPAN R, RAY K. (1997). Reasoning with Vague Default[J]. Fuzzy Sets and Systems, 1997, 91:327-338.

[139] SYAU Y R, HSIEH H T, LEE E S. Fuzzy Numbers in the Credit Rating of Enterprise Financial Condition[J]. Review of Quantitative Finance and Accounting, 2001, 17(4):351.

[140] TANAKA H, UEJIMA U, ASAI K. Linear Regression Analysis with Fuzzy Model [J]. IEEE Transactions on Systems, Man and Cybernetics, 1982, 12(6): 903-907.

[141] TANAKA H, ISHIBUCHI H. An Architecture of Neural Networks with Interval Weights and Its Application to Fuzzy Regression Analysis[J]. Fuzzy Sets and Systems, 1993, 57:27-39.

[142] TSENG F M, TZENG G H, YU H C. Fuzzy Seasonal Time Series for Forecasting the Production Value of the Mechanical Industry in Taiwan[J]. Technological Forecasting and Social Change, 1999, 60 (3): 263-273.

[143] TSENG F M, TZENG G H, YU H C, et al. Fuzzy ARIMA Model for Forecasting the Foreign Exchange Market[J]. Fuzzy Sets and Systems, 2001, 118:9-19.

[144] TSENG T, KLEIN C. A new Algorithm for Fuzzy Multicriteria Decision Making [J]. International Journal of Approximate Reasoning, 1992, 6:45-66.

[145] TSYBOKOV A B. On Nonparametric Estimation of Density Level Sets[J]. Ann. Statist, 1997, 25(3):948-969.

[146] TURNQUIST M A, NOZICK L K. A Nonlinear Optimization Model of Time and Resource Allocation Decisions in Projects With Uncertainty[J]. Engineering Management Journal, 2004, 16(1):40.

[147] UEMURA Y. A Decision Rule on Fuzzy Events[J]. Japanese J. Fuzzy Theory and Systems, 1991, 3:291-300.

[148] UEMURA Y. A Decision Rule on Fuzzy Events under an Observation[J]. J. of Fuzzy Math, 1993, 1:39-52.

[149] UHRMACHER A. Reasoning about Changing Structure: A Modeling Concept for

Ecological Systems[J]. Applied Artificial Intelligence, 1995, 9:157-180.

[150] VAPNIK V N. The Nature of Statistical Learning Theory [M]. 2nd ed. New York: Springer, 2000.

[151] WANG H F, TSAUR R C. Resolution of Fuzzy Regression Model[J]. European Journal of Operational Research, 2000, 126(3):637.

[152] WANG L H, WU S, SHIH H J, et al. A Fuzzy Analysis of Systematic Risk under Price Limits:The Case of the Taiwan Stock Market[J]. International Journal of Management, 2000, 17(4):435.

[153] WANG T Y, CHEN L H. Mean Shifts Detection and Classification in Multivariate Process: A Neural-fuzzy Approach[J]. Journal of Intelligent Manufacturing, 2000, 13(3):211.

[154] WANG C H. Predicting Tourism Demand Using Fuzzy Time Series and Hybrid Grey Theory[J]. Tourism Management, 2004, 25(3):367-374.

[155] WANG H, BELL P. Fuzzy Clustering Analysis and Multifactorial Evaluation for Students' Imaginative Power in Physics Problem Solving[J]. Fuzzy Sets and Systems, 1996, 78:95-105.

[156] WATANABE N, IMAIZUMI T. A Fuzzy Statistical Test of Fuzzy Hypotheses[J]. Fuzzy Sets and Systems, 1993, 53:167-178.

[157] WEI C P, LIU O R, HU J H. (2002). Fuzzy Statistics Estimation in Supporting Multidatabase Query Optimization[J]. Electronic Commerce Research, 2002, 2(3):287.

[158] WEICHSELBERGER K. Axiomatic Foundations of the Theory of Interval-probability[J]. Proceedings of the 2-th Gauss Symposium, Conference B: Statistical Science. Munich, 1993, 2(3): 47-64.

[159] WEN F, CHEN L, WU B. Spatial Time Series Analysis in 1997-1998 Eastern Asia Financial Crisis[J]. Asia Pacific Management Review, 2000, 5:331-345.

[160] WONG Z, KLIR G J. Fuzzy Measure Theory[M]. New York: Plenum Press, 1992.

[161] WU B, CHEN M. Use Fuzzy Statistical Methods in Change Periods Detection [J]. Applied Mathematics and Computation, 1996, 99:241-254.

[162] WU B, CHUNG C. Using Genetic Algorithms to Parameters (d, r) Estimation for Threshold AR Models[J]. Computational Statistics and Data Analysis, 2002, 38:315-330.

[163] WU B, YU-YUN H. A New Approach of Bivariate Fuzzy Time Series: With Ap-

plications to the Stock Index Forecasting[J]. International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 2004, 11(6):671-690.

[164] WU B, YU-YUN H. The Use of Kernel Set and Sample Memberships in the Identification of Nonlinear Time Series[J]. Soft Computing Journal, 2004, 8: 207-216.

[165] WU B, HUNG S. A Fuzzy Identification Procedure for Nonlinear Time Series: With Example on ARCH and Bilinear Models[J]. Fuzzy Set and System, 1999, 108:275-287.

[166] WU B, SUN C. Fuzzy Statistics and Computation on the Lexical Semantics[J]. Language, Information and Computation (PACLIC 11), 1996, 108: 337-346.

[167] WU B, SUN C. Interval-Valued Statistics, Fuzzy Logic, and Their Use in Computational Semantics[J]. Journal of Intelligent and Fuzzy Systems, 2001, 11:1-7.

[168] WU B, TSENG N. A New Approach to Fuzzy Regression Models with Application to Business Cycle Analysis[J]. Fuzzy Sets and System, 2002, 130:33-42.

[169] WU B, YANG W. Application of Fuzzy Statistics in the Sampling Survey[J]. Development and Application for the Quantity Methods of Social Science, 1997, 289-316.

[170] WU H C. The Fuzzy Estimators of Fuzzy Parameters Based on Fuzzy Random Variables[J]. European Journal of Operational Research, 2003, 146(1):101-114.

[171] WU H C. Probability Density Functions of Fuzzy Random Variables[J]. Fuzzy Sets and Systems, 1999, 105:139-158.

[172] WU H C. The Law of Large Numbers for Fuzzy Random Variables[J]. Fuzzy Sets and Systems, 2000, 116:245-262.

[173] XU B, DALE D S, HUANG Y, et al. Cotton Color Classification by Fuzzy Logic [J]. Textile Research Journal, 2002, 72(6):504-509.

[174] XU Z W, KHOSHGOFTAAR T M. Software Quality Prediction for High-Assurance Network Telecommunications Systems[J]. The Computer Journal, 2001, 44 (6):557.

[175] YAGER R, FILEV D. Approximate Clustering via the Mountain Method[J]. IEEE Trans. Syst., Man, Cybern, 1994, 24:1279-1284.

[176] YANG M. A Survey of Fuzzy Clustering[J]. Math. Compu. Modelling, 1993, 18(11):1-16.

[177] YANG M, KO C. On Cluster-wise Fuzzy Regression Analysis[J]. IEEE Trans. Systems Man Cybernet, 1997, 27:1-13.

[178] YEUNG D S, TSANG E C C. Weighted Fuzzy Production Rules[J]. Fuzzy Sets and Systems, 1997, 88:299-313.

[179] YOSHINARI Y, PEDRYCZ W, HIROTA K. Construction of Fuzzy Models through Clustering Techniques[J]. Fuzzy Sets and Systems, 1993, 54:157-165.

[180] YUN K K. The Strong Law of Large Numbers for Fuzzy Random Variables[J]. Fuzzy Sets and Systems, 2000, 111:319-323.

[181] ZADEH L A. Fuzzy Sets[J]. Information and Control, 1965, 8:338-353.

[182] ZADEH L A. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning I[J]. Information Science, 1975, 8(3):199-251.

[183] ZADEH L A. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning II[J]. Information Science, 1975, 8(4):301-359.

[184] ZADEH L A. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning III[J]. Information Science, 1975, 9(1):43-80.

[185] ZADEH L A. Towards a Theory of Fuzzy systems. In R. E. Kalman and N. Declaris, eds. [J]. Aspects of Networks and Systems Theory, 1971, 9(1):469-490.

[186] ZADEH L A. Fuzzy Sets as a Basis for a Theory of Possibility[J]. Fuzzy Set and Systems, 1999, 100:9-34.

[187] ZADEH L A. Toward a Generalized Theory of Uncertainty - an Outline[J]. Information Sciences, 2005, 172:1-40.

[188] ZADEH L A. Generalized Theory of Uncertainty (GTU) - Principal Concepts and Ideas[J]. Computational Statistics & Data Analysis, 2006, 51:15-46.

[189] ZENOVICH S V. Integration of Classification Structures[J]. Nauchno - Tekhnicheskaya Informatsiya, Seriya 2, 1994, 28(6):8-12.

[190] ZHANG L J, LIU C M, DAVIS C J, et al. Fuzzy Classification of Ecological Habitats from FIA Data[J]. Forest Science, 2004, 50(1):117-127.

[191] ZIMMERMANN H J. Fuzzy Set Theory and Its Applications[M]. 3rd ed. Dordrecht:Kluwer Academic, 1991.